Abortion bans disrupt faculty hiring and retention p. 115

CRISPR wood for sustainable timber pp. 124 & 216

Multichip quantum networking p. 221

Science

$15
14 JULY 2023
SPECIAL ISSUE
science.org

AAAS

AI

A MACHINE-INTELLIGENT WORLD

SPECIAL SECTION

EDITORIAL

AI safety on whose terms? p. 138

ESSAY

Art in the age of artificial intelligence p. 139

PERSPECTIVES

Artificial intelligence meets medical robotics p. 141

Improving artificial intelligence with games p. 147

Implications of predicting race variables from medical images p. 149

Mitigating bias in AI at the point of care p. 150

Using machine learning to decode animal communication p. 152

POLICY FORUMS

Do we want less automation? p. 155

Generative AI meets copyright p. 158

NEXTGEN VOICES

AI in search of human help p. 162

REVIEWS

Leveraging artificial intelligence in the fight against infectious diseases p. 164

The central role of density functional theory in the AI age p. 170

RELATED ITEMS

EXPERT VOICES COLUMN BY M. MITCHELL 10.1126/SCIENCE.ADJ5957

NEWS FEATURE p. 121

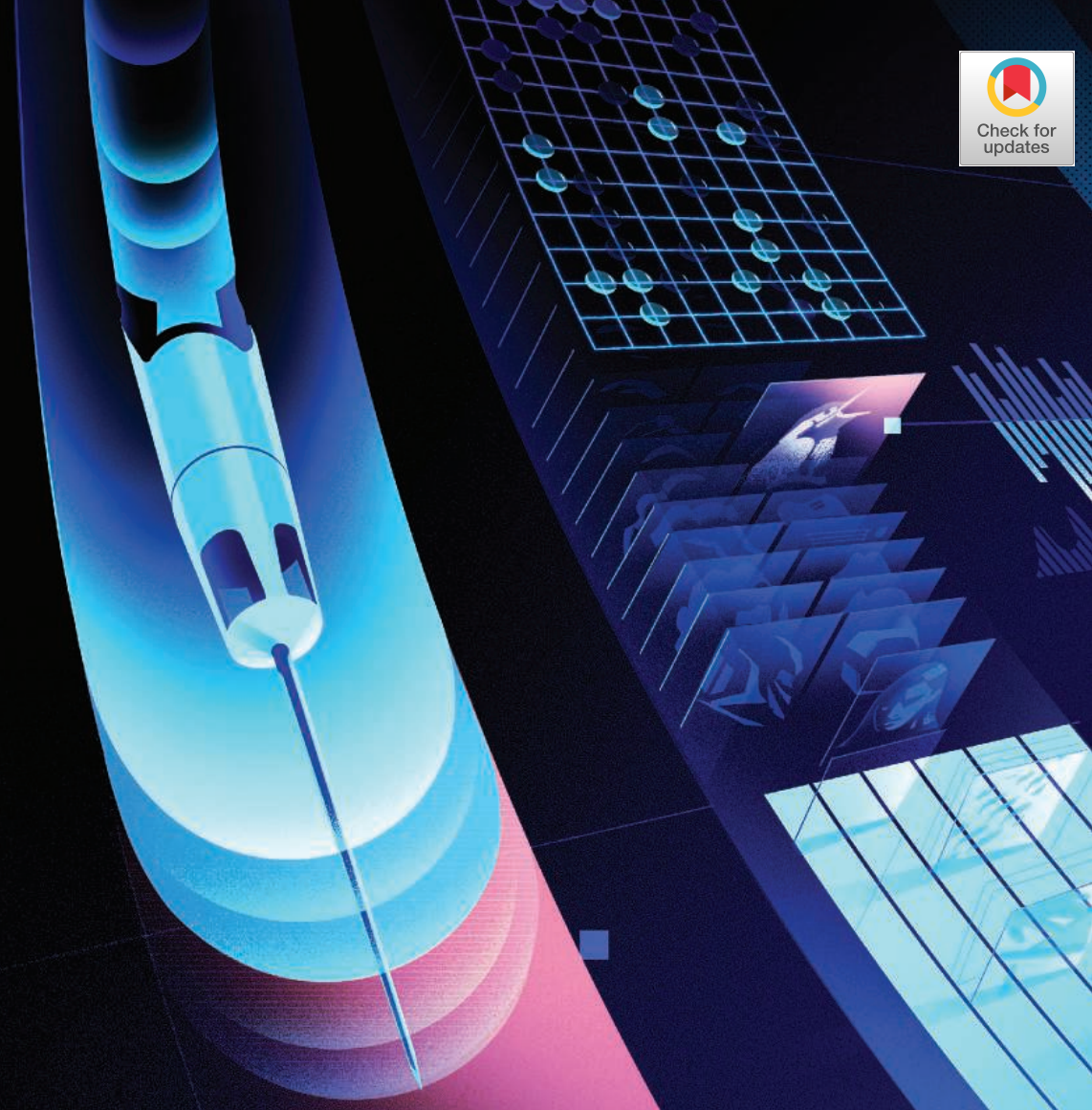
RESEARCH ARTICLE p. 187

SCIENCE ROBOTICS FOCUS BY E. BROADBENT *ET AL.* 10.1126/SCIROBOTICS.ADI6347

SCIENCE ROBOTICS VIEWPOINT BY L. CHRISTOV-MOORE *ET AL.* 10.1126/SCIROBOTICS.ABQ3658

VIDEO

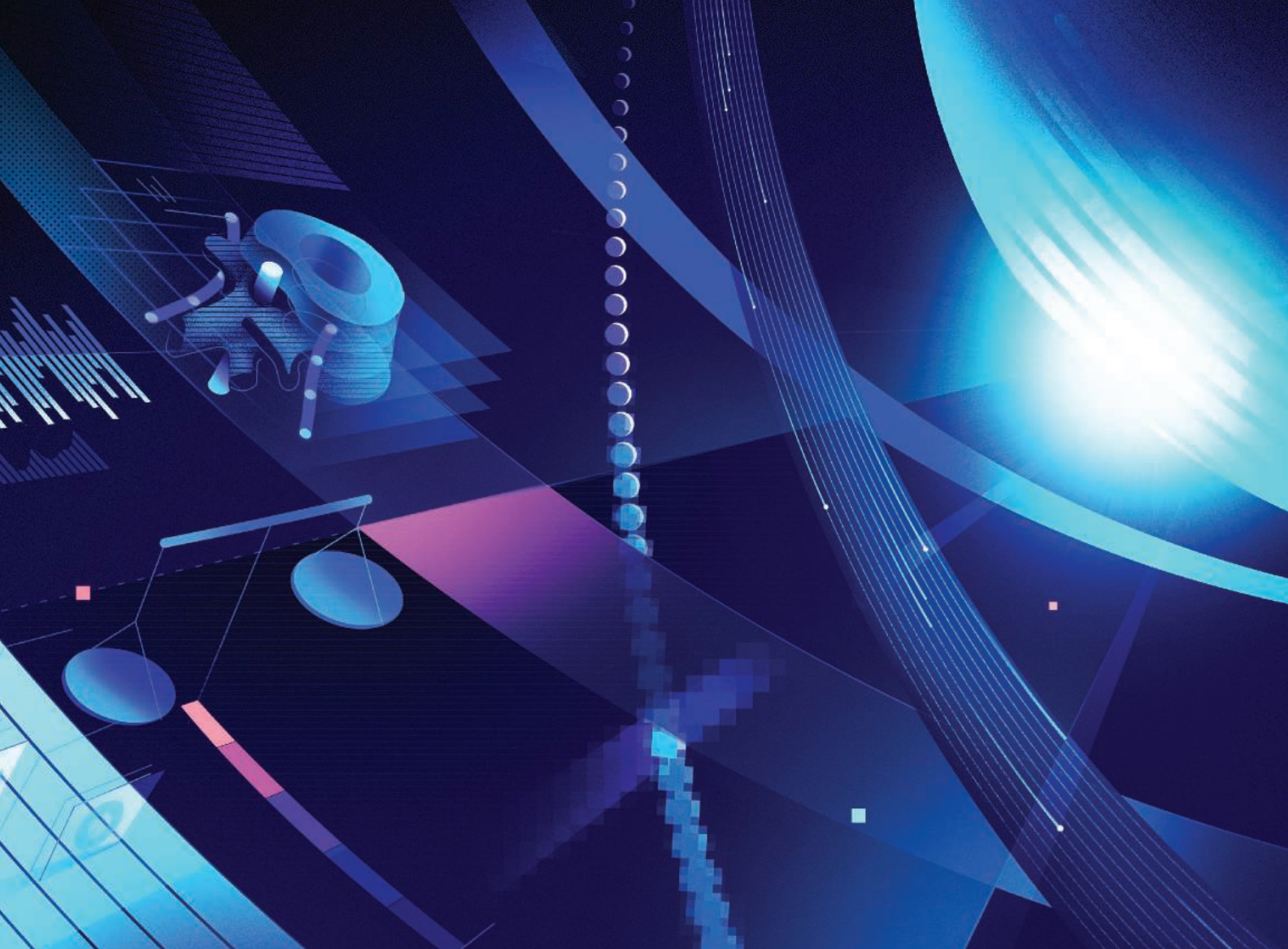
PODCAST



A MACHINE-INTELLIGENT WORLD

Finding a place for AI in science and society

ILLUSTRATION: JASON SOLO/THE JACKY WINTER GROUP



Huge strides have been made in the development of machine-learning algorithms to generate what is commonly called artificial intelligence (AI). Looking to the forefront of how AI is being used in science and society reveals many benefits, as well as grand challenges, that must be addressed.

AI offers considerable benefits, particularly when applications involve substantial amounts of data and high variability. For example, machine learning is being used to facilitate understanding of animal communication, which aims to aid conservation and husbandry. Applications in drug development and chemistry could substantially increase discovery and implementation. Moreover, there has been widespread development and deployment of machine-learning algorithms in medicine, in the form of medical robots to improve patient treatment and recovery and as platforms that facilitate diagnoses and treatment recommendations. But there are concerns about bias and discrimination. Finding ways to develop machine learning without bias and developing AI models that can assess bias and health equity could help maximize the benefits for all.

Such checks and balances are also needed as generative AI is deployed. Is an artifact still "art" if produced by a machine? And what of copyright protection as generative algorithms draw from the work of others? There are also questions about retaining employment and the value of human skills built over years—might those who are most successful in the future be well practiced at using AI models, rather than having a traditional skill honed over years? Concerns about the existential threat of AI systems, to ensure that they are safe for democracy, social justice, and human rights, also need to be tackled and in a transparent manner.

What will a machine-intelligent world of the future look like? AI cannot do everything—yet. It cannot smell, it cannot work in a team, it cannot feel emotions, it cannot invent, and it cannot reason. But it can help us reach further. The question is, how can we ensure that AI is safe so that the benefits outweigh the risks?

10.1126/science.adj4843

EDITORIAL

AI safety on whose terms?

Seth Lazar
is a professor of Philosophy at the Australian National University, Canberra, Australia. seth.lazar@anu.edu.au

Alondra Nelson
is Harold F. Linder Professor in the School of Social Science at the Institute of Advanced Study, Princeton, NJ, USA. anelson@ias.edu

Rapid, widespread adoption of the latest large language models has sparked both excitement and concern about advanced artificial intelligence (AI). In response, many are looking to the field of AI safety for answers. Major AI companies are purportedly investing heavily in this young research program, even as they cut "trust and safety" teams addressing harms from current systems. Governments are taking notice too. The United Kingdom just invested £100 million in a new "Foundation Model Taskforce" and plans an AI safety summit this year. And yet, as research priorities are being set, it is already clear that the prevailing technical agenda for AI safety is inadequate to address critical questions. Only a sociotechnical approach can truly limit current and potential dangers of advanced AI.

Why safety? One could view the shift to safety with cynicism. Big Tech, weary from bad publicity, is seizing the chance to be viewed as saviors from algorithmic harms, not perpetrators of them. Sociotechnical approaches recognize and reject "safety-washing"—giving lip service to safe AI systems, without requisite commitments and practices to ensure this is the case—and call for transparency and accountability to keep companies honest.

"We are making a familiar error."

What does it mean to make AI systems safe, and what values and approaches must be applied to do so? Is it about "alignment," ensuring that deployment of AI complies with some designers' intent? Or is it solely about preventing the destruction of humanity by advanced AI? These goals are clearly insufficient. An AI system capable of annihilating humankind, even if we managed to prevent it from doing so, would still be among the most powerful technologies ever created and would need to abide by a much richer set of values and intentions. And long before such powerful "rogue" AI systems are built, many others will be made that people will use dangerously in their self-interest. Years of sociotechnical research show that advanced digital technologies, left unchecked, are used to pursue power and profit at the expense of human rights, social justice, and democracy. Making advanced AI safe means understanding and mitigating risks to those values, too. And a sociotechnical approach emphasizes that no group of experts (especially not technologists alone) should unilaterally decide what risks count, what harms matter, and to which values safe AI should be aligned. Making AI safe will require urgent public debate on all of these questions and on whether we should be trying to build so-called "god-like" AI systems at all.

How should AI systems be made safe? Narrowly technical approaches are not enough. Instead, AI's use must be targeted in the context of broader sociotechnical systems in which it is always embedded. This means considering the political economy of AI and recognizing when over-indexing on hypothetical future harms predictably risks entrenching the power of a few leading companies, leaving major current challenges unaddressed. It means emphasizing the scarce environmental resources, expropriated data, and exploited labor used to make advanced AI systems. It means considering how advanced AI will be used—for example, to further authoritarianism and illiberalism through mass surveillance and manipulation—which necessitates understanding people and societies, not just their technologies. And it means prioritizing empirically grounded work on actual AI systems and the risks they pose, rather than the (perhaps impossible) task of designing technical mitigations for risks from systems that do not yet exist.

Safety on whose terms? The field of narrow technical AI safety lacks ideological and demographic diversity; it is a near-monoculture and therefore inadequate to the intellectual breadth and rigor required of its mission. Its practitioners, disproportionately white, male, and advantaged, are often drawn from the Silicon Valley social movements of "Effective Altruism" and "Rationalism." A sociotechnical approach resists these structural imbalances, creating spaces for the wider participation necessary to steer our technological future on the basis of equal concern and common humanity. As well as a moral imperative, this is essential for understanding AI risks and building effective solution sets. Women and people of color researchers and advocates have led the way in both revealing the harms of new technologies and in mitigating them; their exclusion from AI safety delegitimizes that field. Sociotechnical AI safety shifts power away from the monoculture.

We are making a familiar error. Faced with disorienting technological change, people instinctively turn to technologists for solutions. But the impacts of advanced AI cannot be mitigated through technical means alone; solutions that do not include broader societal insight will only compound AI's dangers. To really be safe, society needs a sociotechnical approach to AI safety.

–Seth Lazar and Alondra Nelson

10.1126/science.adi8982

Refik Anadol's *Unsupervised* creates imagery inspired by works at the Museum of Modern Art that mesmerizes viewers.

CREDIT: © 2023 THE MUSEUM OF MODERN ART; PHOTO: ROBERT GERHARDT

ESSAY

Art in the age of artificial intelligence

There is more to art than AI-created artifacts, but computational creativity is worth pursuing

By **Joanna Zylinska**

After being expelled from the Royal Academy of Fine Arts in Madrid just before graduation in 1926 for insulting the examination panel, 22-year-old Salvador Dalí headed for Paris. This change of location proved transformative, both for the young artist and for art itself. In France, Dalí went from imitating his masters to developing his own distinctive style. Merging imagery inspired by Sigmund Freud's interpretations of erotic dreams with surreal hallucinations, Dalí embraced association and automaticity as his creative method. This shift led to an exuberant production of canvases and other artifacts, an energetic pace that he maintained until his death in 1989. It is therefore highly apt that when, in 2021, OpenAI launched its deep-learning model capable of generating novel images from natural-language descriptions, it called the program "DALL-E."

A tribute to the famous surrealist artist as well as to Pixar's animated robot WALL-E, the current version of the model—known as DALL-E 2—features automaticity, generativity, and a dreamy aesthetic. Made widely available, DALL-E 2 and its fellow image generators Midjourney and Stable Diffusion have evoked fascination and horror in both the art community and the general public. And understandably so, given that such models have the potential to be more transformational than even Dalí.

Artificial intelligence (AI)–based image generators have raised the prospect of the emergence not just of a radically new art movement but of the end of art itself—and especially of the artist as individual human genius. Their arrival has also posed a fundamental philosophical question: Can machines be truly creative (*1*)? Salvador Dalí's artistic trajectory can help us find some answers, illustrating rather poignantly the key issues that underpin AI art, or, to give the phenomenon its proper name, multisensory computational creativity enabled by machine learning (ML).

CODIFYING CREATIVITY

One precursor to Dalí's artistic shift was his adoption of a "style transfer" through which he painted first like a Renaissance master and then like Picasso. This same imitative method is presently employed in projects such as Microsoft's *The Next Rembrandt*.

Throughout his career, Dalí mined existing repositories of art history resources and used what he found there as inspiration to produce strikingly new artifacts. In effect, the Spanish artist adopted a "combinatorial" method, remixing earlier styles and tropes to arrive at something that looked truly original to his contemporaries. With this approach, Dalí put into practice an understanding of creativity as "the ability to come up with ideas or artefacts that are *new*, *surprising*, and *valuable*" that was codified nearly a century later by computer scientist Margaret Boden in her celebrated book *Creativity and Art* (*2*).

Boden's definition is itself novel because it allows for the possibility that creativity may be exercised by both humans and machines, thus opening the door to defining art in a way that exceeds its humanist legacy. Her understanding of creativity can be applied to both Dalí and DALL-E if we recognize humans' continuous evolution with and through technologies, from stone tools and weapons, through industrial machines, to digital networks—an idea articulated by philosopher Bernard Stiegler (*3*).

It is worth remembering that the anxiety evoked by ML technology in relation to

The author teaches media philosophy and critical digital practice at the Department of Digital Humanities, King's College London, London WC2R 2LS, UK. Email: joanna.zylinska@kcl.ac.uk

art has historical precedence. In the early 1820s, for example, it was feared that the invention of photography would lead to the death of painting. Instead, photography generated an explosion of new ways to see and create images—including painted ones.

AVERAGED DATA, AVERAGE ARTIFACTS

Novelty and exuberance aside, the first phase of AI art enabled by ML was premised on an engineering-based approach to creativity—and on quite a conservative idea of art. Art in this context amounted to little more than the averaging of possibilities, as evidenced in the works of computer science–trained creators such as Mike Tyka, Mario Klingemann, and Memo Akten.

Deploying generative adversarial network (GAN) technology, their images explored the visual effects that arise when an ML model is trained on a database of photographs of human faces or historical artworks and made to produce their approximations. Divergences from the source material, obtained through manipulating the learning algorithm and an inevitable element of surprise, led models to create hallucinogenic images of still or morphing faces in the style of Francis Bacon, with a sprinkling of Dalí's "melting clock" aesthetic. The result was thus not just the averaging of data but also the production of rather average artifacts, albeit ones enveloped in narratives about AI's supposed revolutionary potential.

How should we think about the qualities of these new artifacts deemed "art," and by whom should judgments about them be made? First, we must recognize that there is a history to art that must be approached critically. Art-making is not merely the production of "beautiful" artifacts; the very notion of beauty is itself historical and contextual. In addition, art has recipients to whom its various meanings matter, contested as they may be. In some iterations, art also has financial value, is treated as an investment, and functions as part of a global exchange of capital.

PERFORMING CREATIVE POSSIBILITIES

Artist Refik Anadol's *Unsupervised* (2021–2023) takes the GAN aesthetic and logic to an extreme—but it also opens up some intriguing visual and conceptual avenues. For the project, Anadol trained an ML model on a database of works from the Museum of Modern Art's collection, which includes pieces created over a span of 200 years. The resulting display, shown on a huge vertical screen, offers mesmerizing sequences of morphic images that hint at its data source while creating some new connections between, and reworkings of, the original pieces.

Ungenerous readings have compared this project to a screensaver or a lava lamp, but acerbic reviews do not convey the palpable pleasure experienced by audiences standing in front of Anadol's work. Viewers are often visibly affected by the display, moving their bodies with the undulations of the screen images.

Unsupervised manifests deep existential undertones, unwittingly enacting a proposition by the cybernetics-inspired philosopher of technology Vilém Flusser about image-making being always also a function of the apparatus, and not just of the discrete human mind (*4*). Acknowledging that his work is the outcome of a collaboration between human and machine, Anadol explains: "With the same data, we can generate infinite versions of the same sculpture, but choosing this moment, and creating this moment in time and space, is the moment of creation" (*5*).

The surprising amalgamations that appear in *Unsupervised* invite audiences to reflect on a potential role for machine learning in the creative process.

Because of the computational power available today, infinite variations can be enacted for us by machines. The force of Anadol's work therefore lies not only in its mesmerizing visuality but also in its performance of the possibilities within the latent space of the ML model.

AI ART ENABLES REFLECTION

To say that machines can simulate intelligence, identify patterns, generate novel outputs, or cocreate work with humans is not to suggest their sentience or sapience. Rather it is to recognize that what humans see as intelligent behavior, patterns, and outputs is conditioned by both our cortico-corporeal apparatus and by our environment. By returning our values and meanings back to us for further reflection, machines can help us see ourselves as embodied beings who exist in a sociocultural context.

Many artists working today engage in this kind of second-level inquiry into the conditions of artistic and technological creation with AI. These include studies of algorithmic bias in the works of Jake Elwes and Murad Khan as well as Katja Novitskova's exploration of AI art as "art for another intelligence" (*6*).

POLICY, NOT PANIC, WILL HELP ARTISTS

What lies beneath current questions about machinic creativity is a deeper anxiety about the ways humans can continue to be creative, for how much longer, and at what cost. Given the outpouring of visual, textual, and sonic artifacts enabled by AI, we have begun to ponder the future of human creative professions and pastimes. These developments raise questions about labor, about how we value artifacts and institutions that enable the experience of art, and about art education. Even though artists have always borrowed, copied, and remixed material to produce their wares, the plundering of cultural repositories, without due compensation or, indeed, recognition of individual creators, for the sake of the construction of AI training models needs to be interrogated, for reasons of moral and economic justice, if not humanist panic.

More ominously, however, reducing art to the production of artifacts, in the way hype around "generative AI" has managed to do, could be read as part of a deeper agenda; one concerned with devaluing any form of human cultural activity that cannot be easily discretized and monetized. If we want to live in a world in which there is room for future Dalís and not just new iterations of DALL-E, we need less moral panic about "creative AI" and more sensible policy work around creative education, art institutions, and funding models that ensures such practices are not governed only by the logic of the financial markets or the optimization efforts of Big Tech. ■

REFERENCES AND NOTES

1. For a critical take on this question, see Joanna Zylinska's *AI Art: Machine Visions and Warped Dreams* (Open Humanities Press, 2020).
2. M. A. Boden, *Creativity and Art: Three Roads to Surprise* (Oxford Univ. Press, 2010), p. 29.
3. B. Stiegler, *Technics and Time, 1: The Fault of Epimetheus* (Stanford Univ. Press, 1998).
4. V. Flusser, *Towards a Philosophy of Photography* (Reaktion Books, 2000), pp. 26–27.
5. R. Anadol *et al.*, "Modern dream: How Refik Anadol is using machine learning and NFTs to interpret MoMA's collection," *MoMA Magazine*, 15 November 2021; https://www.moma.org/magazine/articles/658.
6. T. Lykkeberg, "Art for another intelligence" in *If Only You Could See What I've Seen with Your Eyes*, K. Ilves, K. Novitskova, Eds. (Sternberg Press, 2017), pp. 28–40.

10.1126/science.adh0575

PERSPECTIVE

Artificial intelligence meets medical robotics

Artificial intelligence (AI) applications in medical robots are bringing a new era to medicine. Advanced medical robots can perform diagnostic and surgical procedures, aid rehabilitation, and provide symbiotic prosthetics to replace limbs. The technology used in these devices, including computer vision, medical image analysis, haptics, navigation, precise manipulation, and machine learning (ML), could allow autonomous robots to carry out diagnostic imaging, remote surgery, surgical subtasks, or even entire surgical procedures. Moreover, AI in rehabilitation devices and advanced prosthetics can provide individualized support, as well as improved functionality and mobility (see the figure). The combination of extraordinary advances in robotics, medicine, materials science, and computing could bring safer, more efficient, and more widely available patient care in the future. **–Gemma K. Alderton**

Leveraging AI for medical image–guided robotics

By **Michael Yip**[1] *and* **Septimiu Salcudean**[2]

Medical image–guided robotics combines medical images, where locations of key anatomy, lesions, and objects can be identified, and robotics, where the precision placement of instruments or tools provides substantial advantages. Frequently leveraged imaging modalities include ultrasound, magnetic resonance imaging, computed tomography, and white-light or fluorescent endoscopy. The robot may be used to assist in imaging anatomy, or the imaging may assist the robot in guiding it to key targets.

Many early use cases for AI in medical image–guided robotics involved steering instruments, typically needles, to identify anatomy for biopsies. The AI focus was typically on the steering mechanics and planning algorithms because traversing soft tissue involved curvilinear paths that were constrained by the minimum radius of curvature through the tissue, as well as tissue displacements during instrument insertion. Although challenges persist with solving navigation for minimally invasive robotic tools with multiple and intermittent contacts with tissue, a large scope of solutions involving AI-driven planning for robotic steering of instruments are now available (*1*, *2*), and highly dexterous robotic systems have been proposed that can plan and reach targets with the aid of AI (*3*).

Most effort now concerns image understanding. Previous techniques used tedious, human-segmented (hand-annotated) anatomical features and/or weak, feature-based recognition of anatomy of interest. New AI strategies for image guidance leverages semantic information—i.e., higher-level reasoning about the type of anatomy and its characteristics, identified directly from pixel information, to provide improved, safer navigation. As with the considerable improvements in object localization and scene segmentation in computer vision, these techniques have begun their translation to surgical scenes and surgically relevant objects (*4*). These translated techniques can then be leveraged by robots to plan and reach targets with high accuracy.

An interesting use of image understanding is in image acquisition itself (*5*). For example, robot assistance makes medical ultrasound more consistent, enables autonomous scanning, and may improve access to examinations in remote and underserved communities. Moving the ultrasound transducer against the patient to identify the standard imaging planes used in medical diagnostics is a challenging AI problem involving deep learning networks to predict necessary probe motions (*6*), reinforcement learning (RL), and learning by demonstration (*7*). Robotic assistance is also used in endoscopy, where rigid or flexible endoscopes, and even capsules with magnetic actuation, are deployed in surgery, gastrointestinal tract imaging, and bronchoscopy (*8*). Maneuvering endoscopes is challenging and requires considerable experience to be mastered, making automatic motions attractive. Approaches based on learning by demonstration are promising, e.g., by using behavioral cloning (*6*). AI techniques are used in endoscope localization and mapping and in the analysis of the very large image datasets involved in diagnosis (*9*).

Intraoperative ultrasound and x-ray imaging enable the registration of preoperative medical images to the patients for biopsy and surgery. Localizing targets identified through intraoperative registration in the corresponding laparoscopic camera views remains a challenge. Maintaining such registration requires long-term tissue tracking using computer vision techniques, in the presence of tissue changes as the intervention progresses (*4*).

Unlike building neural network models for everyday scenes, finding labeled data for training models for medical robotic applications is a substantial bottleneck. Accurate labeling must be done by trained professionals such as surgeons and radiologists. Thus, getting ground-truth data is very expensive and not scalable. Synthetically generated images help to address a part of this problem, but synthetic images are sufficiently different from real images and can lead to overfitting (*10*). Given the related challenge of human labeling of intraoperative video sequences to train neural networks, unsupervised or weakly supervised approaches are desirable (*11*).

[1]Electrical and Computer Engineering, University of California San Diego, La Jolla, CA, USA. [2]Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. yip@ucsd.edu; tims@ece.ubc.ca [3]Department of Industrial Engineering and Operations Research and Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA, USA. goldberg@berkeley.edu [4]School of Engineering and Materials Science, Queen Mary University of London, London, UK. [5]The BioRobotics Institute, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, Pisa, Italy. arianna.menciassi@santannapisa.it; k.althoefer@qmul.ac.uk [6]Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA. [7]Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA. [8]Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA. axel@jhu.edu [9]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Boston, MA, USA. [10]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Boston, MA, USA. krithika@seas.harvard.edu; walsh@seas.harvard.edu [11]Joint Department of Biomedical Engineering, North Carolina State University, Raleigh, NC, USA. [12]Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. hhuang11@ncsu.edu

Smart medical robots

Various applications of artificial intelligence (AI), including machine learning, machine vision, and haptic control, have resulted in the development of robotic devices that can be used in all aspects of patient care, including diagnostics, surgical procedures, rehabilitation, and limb replacement. The use of robotics in medicine aims to ensure consistent, safe, and efficient treatment, as well as allowing data gathering for improvement and potentially increasing access to treatment in underserved communities and remote regions, and to those affected by natural disasters.

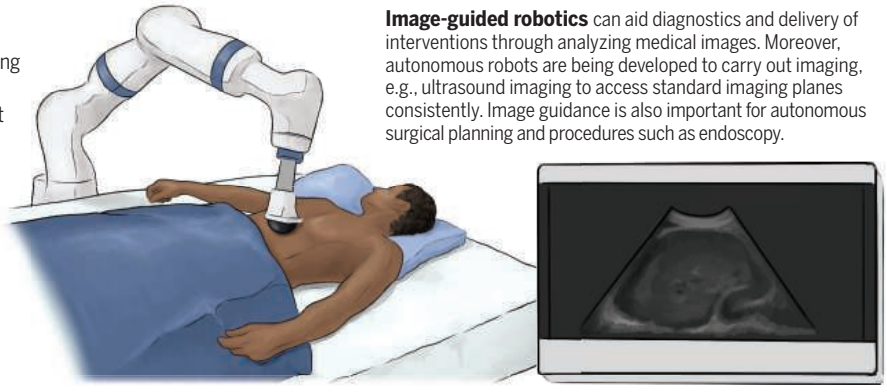
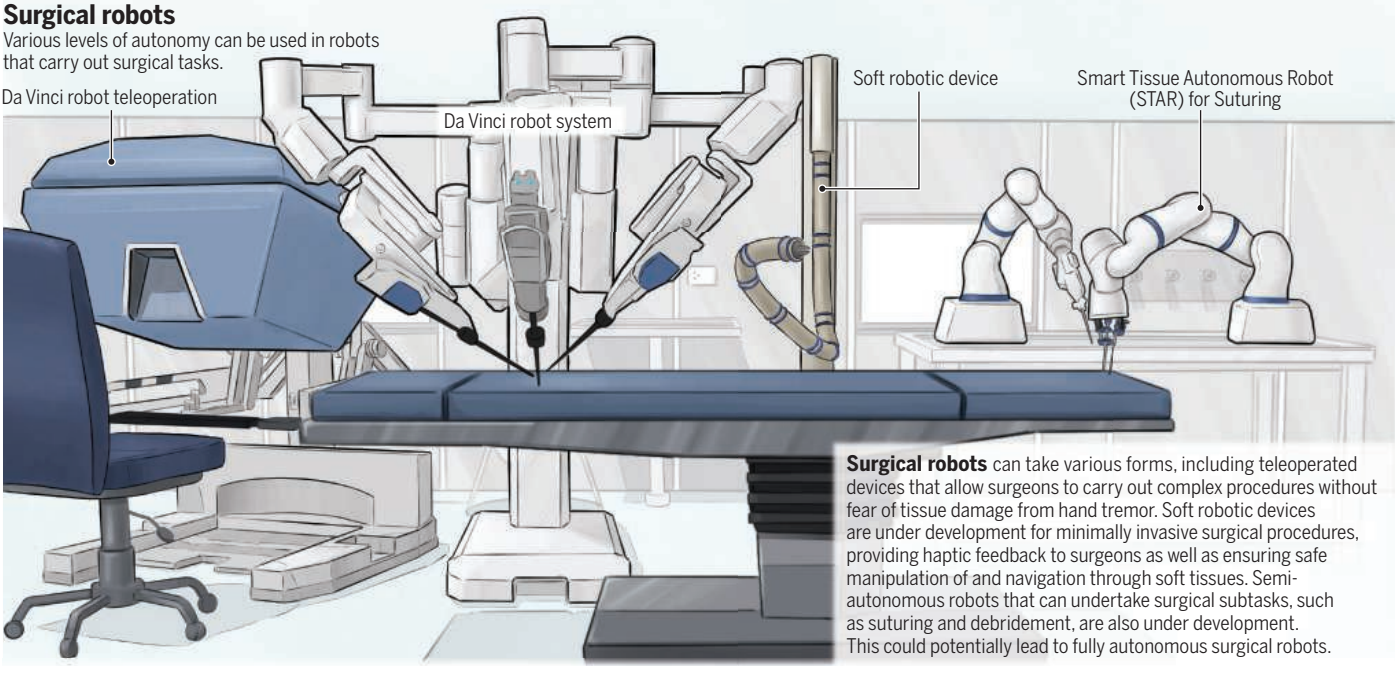


Image-guided robotics can aid diagnostics and delivery of interventions through analyzing medical images. Moreover, autonomous robots are being developed to carry out imaging, e.g., ultrasound imaging to access standard imaging planes consistently. Image guidance is also important for autonomous surgical planning and procedures such as endoscopy.

Surgical robots

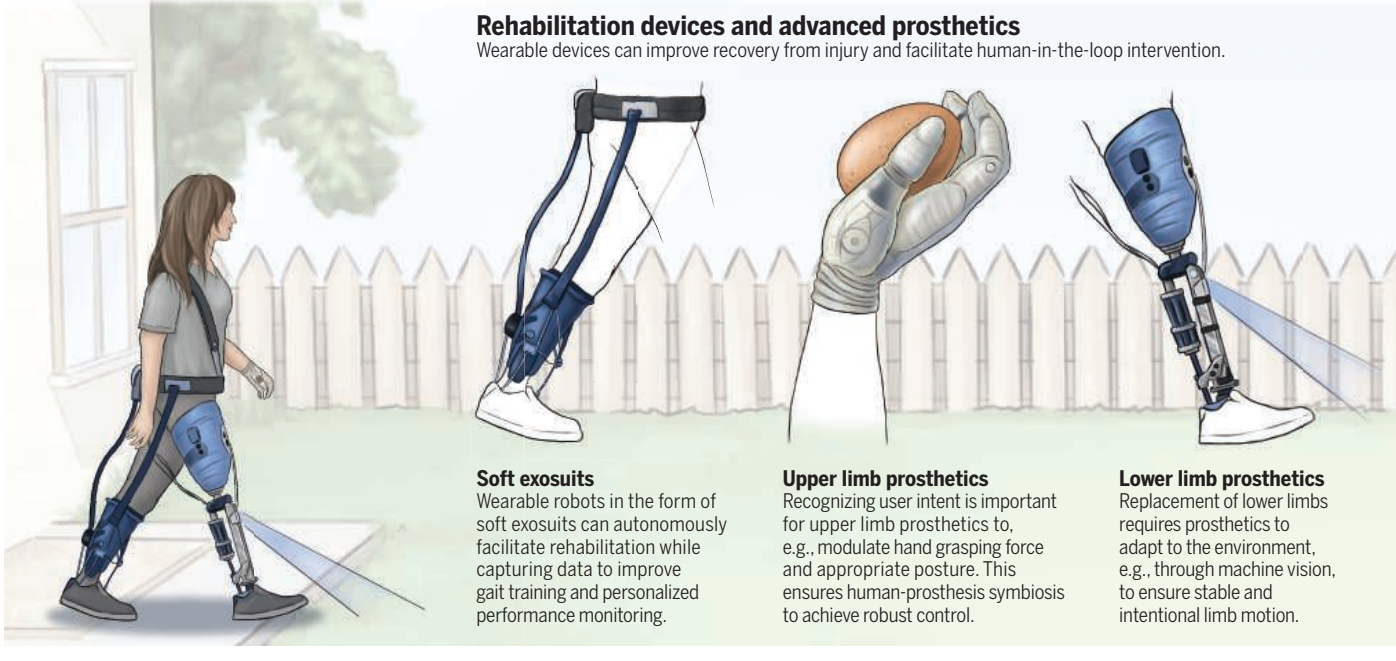
Various levels of autonomy can be used in robots that carry out surgical tasks.


Da Vinci robot teleoperation
Da Vinci robot system
Soft robotic device
Smart Tissue Autonomous Robot (STAR) for Suturing

Surgical robots can take various forms, including teleoperated devices that allow surgeons to carry out complex procedures without fear of tissue damage from hand tremor. Soft robotic devices are under development for minimally invasive surgical procedures, providing haptic feedback to surgeons as well as ensuring safe manipulation of and navigation through soft tissues. Semi-autonomous robots that can undertake surgical subtasks, such as suturing and debridement, are also under development. This could potentially lead to fully autonomous surgical robots.

Rehabilitation devices and advanced prosthetics

Wearable devices can improve recovery from injury and facilitate human-in-the-loop intervention.



Soft exosuits
Wearable robots in the form of soft exosuits can autonomously facilitate rehabilitation while capturing data to improve gait training and personalized performance monitoring.

Upper limb prosthetics
Recognizing user intent is important for upper limb prosthetics to, e.g., modulate hand grasping force and appropriate posture. This ensures human-prosthesis symbiosis to achieve robust control.

Lower limb prosthetics
Replacement of lower limbs requires prosthetics to adapt to the environment, e.g., through machine vision, to ensure stable and intentional limb motion.

Ultimately, the combination of new techniques in data-efficient learning, coordinated efforts in growing and retaining labeled medical image datasets, and advances in safe and dexterous robotic systems will push the relevance of medical image–guided robotics to the forefront of the next generation of interventional and diagnostic care.

Supervised autonomy in robot-assisted surgery and telesurgery

By **Ken Goldberg**[3]

Each year, more than a million surgeries are performed with robots (*12*). These robots are very sophisticated but not at all autonomous—they are fully controlled by human surgeons. This is because surgery is fault-intolerant, there are a vast number of rare but potentially dangerous edge conditions, and the consequences of even a single failure can be fatal. It may be a long time before fully autonomous robots are sufficiently safe and reliable for the clinic. In the meantime, substantial progress is being made in the use of "supervised autonomy," where specific subtasks are performed by a robot under close supervision of a human, who is ready to take over when necessary (*13*, *14*).

Consider a surgical subtask such as debridement: the removal of damaged tissues or foreign fragments from a wound. Debridement can be very tedious; it is easy for a surgeon to overlook fragments, which can lead to infections. Debridement is a surgical task that could benefit greatly from supervised autonomy, whereby a surgical robot and camera system could systematically identify and remove fragments under close supervision of the surgeon who is ready to take over if the system misidentifies a miscolored human tissue as a foreign fragment. Supervised autonomy for surgical debridement has been achieved in laboratory conditions (*15*, *16*), but research is still needed to extend these results and evaluate them in vivo.

Another example is surgical suturing. This surgical subtask is often left to medical residents because it is tedious and somewhat fault-tolerant. Surgical suturing requires even placement of sutures that balance tissue forces. Supervised autonomy could produce sutures that are more consistent, thus reducing healing time and scarring.

Researchers are studying how suturing could be performed using supervised autonomy, in which the surgeon can outline a wound by touching its boundaries with an instrument, and the system computes and displays an optimal placement of needle entry and exit points that evenly distributes tension across the wound. The surgeon can then adjust these points or allow the robot to autonomously perform the suturing under close supervision. Researchers have demonstrated initial results in the laboratory (*17*, *18*), but securely gripping surgical needles during insertion, handing needles back and forth between gripper tools, and managing the slack in surgical thread must be addressed before suturing can be considered for testing in the clinic.

Supervised autonomy also opens the door to "telesurgery," whereby an experienced surgical expert can guide a surgery at a distant location. Telesurgery has the potential to considerably increase access to skilled surgeons in remote regions or during a natural disaster (*19*). Direct control of all surgical instrument motions by the expert is not possible because of the inherent time delays of optical and electrical signals, which cause any direct control loop to be unstable. Supervised autonomy can solve this problem by allowing surgical subtasks to be locally controlled, with intermittent remote supervision. Telesurgery has been demonstrated with the use of dedicated high-speed fiber optic networks and is being actively explored by researchers, but it is not yet approved for clinical use.

Advances in sensors and ML have evolved robotics considerably in recent years, and there are many opportunities for AI in the operating room. AI can be used to enhance digital camera images and answer spoken queries from a surgeon during a procedure. But supervised autonomy offers the greatest potential for reducing tedium and improving patient outcomes. Researchers worldwide are exploring how supervised autonomy can be used to enhance robot-assisted surgery and telesurgery (*20*). As the original patents on surgical robotics expire (*21*), new commercial surgical robot systems are increasing the diversity of hardware and interfaces. This competition is motivating the commercial development of new functionality, and it is likely that supervised autonomy will be available in the clinic in the coming decade.

Soft robotics for minimally invasive surgery

By **Kaspar Althoefer**[4] *and* **Arianna Menciassi**[5]

Over the past few decades, there have been considerable advancements in robot-assisted minimally invasive surgery (RAMIS). RAMIS systems use slender, straight-line instruments to operate through small incisions in the patient's skin. Robotics makes procedures simpler, filtering the manual tremor of surgeons, improving overall ergonomics, and restoring three-dimensional (3D) vision that is normally not possible in manual minimally invasive procedures. In addition, RAMIS allows the generation of a huge amount of data that can be used for improving safety and implementing some autonomous tasks (*22*). Despite the success of some RAMIS platforms, such as the da Vinci Surgical System for prostatectomy and abdominal or thoracic surgery, these systems are often limited by their rigid component design, which can make it difficult to access certain areas of the body and can lead to tissue injuries.

Soft robotics is a promising avenue for developing more flexible and adaptable surgical robots, with the necessary dexterity and stiffness modulation to perform surgical procedures safely. The key feature of soft robotics is the use of materials that can deform, bend, shrink, and change stiffness (*23*), pushing the paradigm of robotic surgery in a safer and softer direction. These robots address different body regions, such as the ear, abdomen, and thorax, and they can be dedicated both for diagnosis and intervention. For example, a fluid-driven soft robotic system was developed for increasing patient comfort during ear therapy and safely steering a needle to the desired injection site (*24*). Diagnosis of gastrointestinal tract pathology is also a key application for soft robots, because these tissues are flexible, stretchable, and often collapsed, requiring a spectrum of soft and stiff working modalities (*25*). A capsule robot for endoscopy that uses eversion navigation and a soft shape-shifting mechanism has been recently demonstrated (*26*).

A large-scale project to explore soft robotics for RAMIS was the European Union project STIFF-FLOP (stiffness controllable flexible and learnable manipulator for surgical operations) from 2012 to 2015 (*27*). The soft robotic systems that were developed were made from biocompatible silicone rubber and pneumatically actuated, by using new fabrication methods that allow for the creation

The Smart Tissue Autonomous Robot (STAR), which is holding a needle driver, needle, and suture in preparation for suturing intestinal tissues, is inspected.

of reliable structures that are also safe and effective. In addition, advanced ML techniques were employed to intuitively teleoperate the soft robots in the abdominal cavity of the patient, and haptic systems allowed surgeons to discern interactions of the robot with the soft tissue environment.

There remain substantial technical challenges (*28*). A major issue is the lack of precision and accuracy in soft robotic systems. In traditional surgical robots, electrical motors are used directly or by means of tendons to move the robot's arms, and effectors are made from rigid components that do not deform during operation. However, soft robotic systems rely on deformation of the material that the robot is constructed from to achieve movement. The resultant motion is more difficult to model and can result in lower positional accuracy, which could be a critical concern in surgery (*29*). To overcome this challenge, advanced strategies based on AI, ML, and data-driven control that can cope with the highly nonlinear motion behavior of soft robots are being developed. Recent advances in computer power, computer vision, ML, real-time modeling, and simulation can make operation of soft robots for surgery possible without cumbersome teleoperation modalities and extensive training sessions for surgeons (*30*).

Will soft robots for RAMIS replace well-established surgical robots, or will soft robotic design rules, relying on morphological computation (*31*), permeate traditional technologies for RAMIS? What benefits the patient most needs to drive research in soft robotics surgery.

Bringing highly autonomous surgical robotics to the clinic

By **Justin D. Opfermann**[6,7] *and* **Axel Krieger**[6,7,8]

Autonomous surgical robots are the surgeons of the future. They have the power to standardize patient outcomes independent of a surgeon's experience and skill (*32*), they integrate AI and dexterous tools to perform tasks with more consistency and accuracy than expert surgeons (*33*), and they can provide essential care in environments where no surgeon is available such as human space flights (*34*). Such robots will democratize health care by making quality surgery ubiquitous and by minimizing the incidence of corrective surgery, thereby reducing health care costs. Although most systems are not clinically approved today, they will certainly play a role in the future.

Generally, autonomous surgical robots are classified by their respective level of autonomy (LoA) and incorporate algorithms that are responsible for increased surgical decision-making (*35*). As the level of autonomy increases, so too does the complexity of the robot's role in surgery and the amount of AI integrated with the system. For instance, the LoA 0 (no autonomy) da Vinci Surgical System uses human teleoperation without AI to perform surgery, whereas the LoA 1 (robot assistance) EndoAssist camera

holder uses algorithms to restrict tool motions (*36*). For higher levels of autonomy, the surgeon defers control to the robot, which uses AI to execute surgical tasks. During LoA 2 (task autonomy), a robot may use learning by observation to independently cut tissue as with the da Vinci Research Kit (*37*), whereas the LoA 3 (conditional autonomy) Smart Tissue Autonomous Robot (STAR) uses ML to track soft tissue deformation to execute the surgical plan of suturing (*38*). The LoA 4 (high autonomy) and LoA 5 (full autonomy) surgical systems are not yet feasible with today's technology, but their development is on the horizon.

There remain several technical, regulatory, and social challenges to solve before the highest levels of autonomy can be reached. Robots will need to better detect, process, and respond to unpredictable variations in the surgical field. These challenges will be magnified in soft tissue surgeries, where deep learning and AI will be necessary to predict, and react to, the changing surgical scene (*39*). The resulting technologies will face increased regulatory scrutiny because it is not clear how a system capable of practicing medicine would be regulated by the US Food and Drug Administration (FDA) or medical device regulators elsewhere. These systems will need to demonstrate safety that exceeds teleoperation and efficacy as good as the clinical standard of care. For AI algorithms, this means sensitivity and specificity at least as good as an expert surgeon's. Social pressures might also temper adoption as there is general resistance with using AI in medicine (*40*). To build public trust, autonomous robots will likely follow a stepwise approach to adoption, in which autonomous subtasks such as tissue identification, endoscope control, and suturing will be gradually introduced. These tasks will then be combined into a full procedure, paving the way for autonomous robotic surgery in the operating room of tomorrow.

Rehabilitation robots to go

By **Krithika Swaminathan**[9] *and* **Conor J. Walsh**[10]

Rehabilitation needs to extend from the clinic into the community and home to provide patients with a continuum of care. Toward this goal, engineers, clinicians, and end users have developed wearable robots that allow people with mobility impairments to practice and experience better movement. Although these systems have historically been seen as assistive technologies, ongoing work is demonstrating that these portable and autonomous robots, and the data that they capture, can lead to fundamentally different rehabilitation approaches. One can imagine a future in which wearable robots are used in the clinic to reduce the physical burden on clinicians while learning patient-specific impairments, and are then sent to the patient's home to track their recovery with individualized ML algorithms.

Recent technological advances in hardware and software are facilitating this shift to real-world use (*41*–*44*). Innovative approaches for actuation hardware achieve higher specific powers (power per unit mass) through cable-driven (*45*) or pneumatically-driven mechanisms (*46*). Soft, apparel-based interfaces can be lightweight and nonrestrictive to the user and let a device smoothly transition between assisting motion, being mechanically transparent, and resisting motion for strength training. The increased modularity of these systems also allows for tailoring the technology to individual-specific impairments. Along with new mechanical designs, learning-based estimation algorithms now use wearable sensors to detect and quantify movements (*47*), and control strategies now bring users into the control loop to provide individualized intervention (*48*). These techniques further enable studying the rehabilitative potential of high-dosage and high-intensity training beyond the clinic (*49*).

Wearable cloud-connected robots will usher in the era of data-driven physical telerehabilitation. Integrated sensing can provide clinicians and users with feedback on important biomechanical and physiological metrics, similar to smart fitness trackers. However, optimally combining this feedback with clinical input to guide motor learning and encourage behavioral change remains an open challenge. Moreover, developing generalized data-driven ML and AI algorithms in clinical populations is particularly complex owing to the limited data available from any given individual and the large variability across individuals. Continuous movement assessments also increase the temporal resolution of characterizing a patient's progression, enabling more timely and accurate detection of performance degradation or improvement.

However, a challenge remains in developing estimation algorithms that are robust to noise across recovery timescales (months to years), owing to sensor drift, sensor placement, environmental changes, and day-to-day user variability. Validating these approaches is further complicated by the lack of ground-truth data in real-world settings. If successful, the corresponding data from the patients and robots can be used with ML techniques to identify who is most likely to benefit from a given device (*50*). Such categorization will enable efficient prescription of interventions, minimizing costs for clinicians and patients alike. The field will then have new opportunities to investigate how longitudinal data can inform new methods for individualization of training parameters, such as for biofeedback and device controller optimization.

Data from wearable robots worn in the real-world will inform the design of computational models and experimental validation of human-robot interaction during rehabilitation, ultimately leading to adaptive systems that better synergize with end users.

PHOTO: HARVARD BIODESIGN LAB

A soft robotic, cloud-connected glove facilitates high-dosage and high-intensity self-directed teletherapy for stroke survivors in the home.

AI enables symbiotic robotic prosthetics

By **He (Helen) Huang**[11,12] and **I-Chieh Lee**[11,12]

Advanced robotic prosthetics, such as dexterous prosthetic hands and motorized prosthetic legs, have led to a paradigm shift to restore the mobility of individuals with limb loss (*51*, *52*). These modern prostheses have embedded AI into machine operation to enable adaptation to user intent, environments, and the user's physical condition. This is essential for human-prosthesis symbiosis—an intelligent prosthesis and a human user functioning seamlessly together as one system in daily life (*53*).

For example, AI has enabled neural control of prosthetic limbs. Controlling prosthetic limbs on the basis of user intent from the brain is a fascinating concept. This requires an effective neural decoder that can accurately interpret user intent through human neuromuscular signals for prosthetic limb control. ML algorithms, ranging from simple linear classifiers to deep learning regression models, have been powerful neural decoding methods for recognizing user intent regarding joint motions (e.g., wrist or knee flexion and extension), hand grasping patterns (e.g., fine pinch or power grip), or locomotion modes (e.g., sit-to-stand transition and level ground walking) (*54*, *55*). The ML decoder output is sent to the prosthesis controller to produce the user's intended limb motion to enable human-prosthesis symbiosis during task performance.

Human-prosthesis symbiosis through the implementation of AI allows adaptation to various environments and contexts. Human hands can dexterously interact with objects of different sizes and materials; human legs can adapt to various terrains while walking. Therefore, a symbiotic prosthetic limb should also be environmentally adaptive. Machine vision has been adopted to create environmental awareness for prosthesis control. Through deep learning algorithms applied to images captured by cameras mounted on a prosthetic hand, machine vision can recognize the intended grasping object, which allows the prosthesis arm to prepare the appropriate wrist posture and hand grasping pattern or force to facilitate grasping actions (*56*, *57*). Similarly, vision sensors mounted on prosthetic legs can recognize the terrain in front of the user, which autonomously adapts prosthesis control accordingly for seamless terrain transitions (*58*, *59*).

A symbiotic prosthesis needs to provide personalized assistance to each user owing to the large inter-amputee differences in their physical conditions and motor deficits. In current clinics, personalization of robotic lower-limb prosthesis control is performed manually and heuristically, which is inaccurate and time- and labor-intensive. To automate the process, researchers have developed RL algorithms and other data-driven optimization approaches, such as Bayesian optimization, to tune prosthesis control with the human-in-the-loop for personalized walking assistance (*60*). For RL-based algorithms, the prosthesis personalization procedure can be as short as 5 minutes, and the resulting smart AI tuning agent can continue producing user-adaptive control over different time frames (*61*).

Although AI in robotic prosthetics has shown great promise, AI needs to be more robust and safer for daily prosthesis control owing to having the human-in-the-loop. Additionally, it is an open question whether human users cognitively embody and trust AI-enabled prostheses. These challenges should direct future research efforts toward making AI-enabled symbiotic, robotic prostheses versatile, safe to use, and cognitively acceptable by users with limb amputations.

REFERENCES AND NOTES

1. J. Burgner-Kahrs, D. C. Rucker, H. Choset, *IEEE Trans. Robot.* **31**, 1261 (2015).
2. G. Fichtinger, J. Troccaz, T. Haidegger, *Proc. IEEE* **110**, 932 (2022).
3. D. Schreiber *et al.*, in *2022 International Conference on Robotics and Automation (ICRA)* (IEEE, 2022), pp. 5487–5494.
4. S. Lin *et al.*, in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
5. S. E. Salcudean, H. Moradi, D. G. Black, N. Navab, *Proc. IEEE* **110**, 951 (2022).
6. R. Droste, L. Drukker, A. T. Papageorghiou, A. J. Noble, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 583–592.
7. K. Li, Y. Xu, M. Q. H. Meng, *IEEE Trans. Med. Robot. Bionics* **3**, 510 (2021).
8. Z. Li, P. W. Y. Chiu, *Visc. Med.* **34**, 45 (2018).
9. Z. Fu *et al.*, *IEEE Access* **9**, 41144 (2021).
10. M. J. Willemink *et al.*, *Radiology* **295**, 4 (2020).
11. A. Schmidt, O. Mohareri, S. DiMaio, S. E. Salcudean, in *2022 International Conference on Robotics and Automation (ICRA)* (IEEE, 2022), pp. 1281–1288.
12. M. Tindera, "Robot wars: $60B intuitive surgical dominated its market for 20 years. Now rivals like Alphabet are moving in," *Forbes*, 14 February 2019.
13. T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control* (MIT Press, 1992).
14. C. D'Ettorre *et al.*, *IEEE Robot. Autom. Mag.* **28**, 56 (2021).
15. D. Seita *et al.*, in *IEEE International Conference on Robotics and Automation*, Brisbane, Australia, 21–25 May 2018, pp. 6651–6658.
16. V. Patel *et al.*, in *International Symposium on Medical Robotics (ISMR)*, Atlanta, GA, USA, 1-3 March 2018, pp. 1–6.
17. S. Sen *et al.*, *2016 IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 2016, pp. 4178–4185.
18. Z. Y. Chiu, A. Z. Liao, F. Richter, B. Johnson, M. C. Yip, *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 5286–5292.
19. J. P., R. J. Choi, R. S. Oskouian, R. S. Tubbs, *Cureus* **10**, e2716 (2018).
20. M. Hwang *et al.*, in *IEEE Transactions on Automation Science and Engineering*, 2023, vol. 20, pp. 909–922.
21. C. Kent, "Robotic Surgery: A Race to the Top," *Medical Device Network*, 31 March 2020.
22. P. Fiorini, K. Y. Goldberg, Y. Liu, R. H. Taylor, *Proc. IEEE* **110**, 993 (2022).
23. M. Cianchetti, C. Laschi, A. Menciassi, P. Dario, *Nat. Rev. Mater.* **3**, 143 (2018).
24. L. Lindenroth, S. Bano, A. Stilli, J. G. Manjaly, D. Stoyanov, *IEEE Robot. Autom. Lett.* **6**, 871 (2021).
25. A. Loeve, P. Breedveld, J. Dankelman, *IEEE Pulse* **1**, 26 (2010).
26. V. Consumi, L. Lindenroth, J. Merlin, D. Stoyanov, A. Stilli, *IEEE Robot. Autom. Lett.* **8**, 1659 (2023).
27. M. Cianchetti *et al.*, *Soft Robot.* **1**, 122 (2014).
28. M. Runciman, A. Darzi, G. P. Mylonas, *Soft Robot.* **6**, 423 (2019).
29. K.-W. Kwok, H. Wurdemann, A. Arezzo, A. Menciassi, K. Althoefer, *Proc. IEEE* **110**, 871 (2022).
30. R. H. Taylor, N. Simaan, A. Menciassi, G.-Z. Yang, *Proc. IEEE* **110**, 823 (2022).
31. H. Hauser, T. Nanayakkara, F. Forni, *IEEE Control Syst.* **43**, 114 (2023).
32. S. Tou, M. Gómez Ruiz, A. G. Gallagher, N. J. Eardley, K. E. Matzel, *Colorectal Dis.* **22**, 1826 (2020).
33. M. Hwang *et al.*, *IEEE Trans. Autom. Sci. Eng.* **20**, 909 (2023).
34. D. Pantalone *et al.*, *NPJ Microgravity* **7**, 56 (2021).
35. G. Z. Yang *et al.*, *Sci. Robot.* **2**, eaam8638 (2017).
36. J. M. Gilbert, *Ann. R. Coll. Surg. Engl.* **91**, 389 (2009).
37. A. Murali *et al.*, 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 2015, pp. 1202–1209.
38. H. Saeidi *et al.*, *Sci. Robot.* **7**, eabj2908 (2022).
39. J. Lu, A. Jayakumari, F. Richter, Y. Li, M. C. Yip, *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021, pp. 4783–4789.
40. C. Longoni, A. Bonezzi, C. K. Morewedge, *J. Consum. Res.* **46**, 629 (2019).
41. C. Siviy *et al.*, *Nat. Biomed. Eng.* **7**, 456 (2023).
42. L. M. Mooney, E. J. Rouse, H. M. Herr, *J. Neuroeng. Rehabil.* **11**, 80 (2014).
43. L. N. Awad, A. Esquenazi, G. E. Francisco, K. J. Nolan, A. Jayaraman, *J. Neuroeng. Rehabil.* **17**, 80 (2020).
44. G. S. Sawicki, O. N. Beck, I. Kang, A. J. Young, *J. Neuroeng. Rehabil.* **17**, 25 (2020).
45. L. N. Awad *et al.*, *Sci. Transl. Med.* **9**, eaai9084 (2017).
46. T. Proietti *et al.*, *Sci. Transl. Med.* **15**, eadd1504 (2023).
47. C. P. Adans-Dester, C. E. Lang, D. J. Reinkensmeyer, P. Bonato, in *Neurorehabilitation Technology*, D. J. Reinkensmeyer, L. Marchal-Crespo, V. Dietz, Eds. (Springer, 2022).
48. P. Slade, M. J. Kochenderfer, S. L. Delp, S. H. Collins, *Nature* **610**, 277 (2022).
49. R. W. Nuckols *et al.*, *Ann. N Y Acad. Sci.* 10.1111/nyas.14998 (2023).
50. D. J. Reinkensmeyer *et al.*, *J. Neuroeng. Rehabil.* **13**, 42 (2016).
51. M. Goldfarb, B. E. Lawson, A. H. Shultz, *Sci. Transl. Med.* **5**, 210ps215 (2013).
52. M. Laffranchi *et al.*, *Sci. Robot.* **5**, eabb0467 (2020).
53. H. H. Huang, J. Si, A. Brandt, M. Li, *Curr. Opin. Biomed. Eng.* **20**, 100314 (2021).
54. D. Farina *et al.*, *Nat. Biomed. Eng.* **7**, 473 (2023).
55. A. Fleming *et al.*, *J. Neural Eng.* **18**, 041004 (2021).
56. D. P. McMullen *et al.*, *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**, 784 (2014).
57. M. Markovic, S. Dosen, D. Popovic, B. Graimann, D. Farina, *J. Neural Eng.* **12**, 066022 (2015).
58. K. Zhang *et al.*, *IEEE Trans. Cybern.* **51**, 3285 (2021).
59. B. Zhong, R. L. Da Silva, M. Li, H. Huang, E. Lobaton, *IEEE Trans. Autom. Sci. Eng.* **18**, 458 (2021).
60. R. Gehlhar, M. Tucker, A. J. Young, A. D. Ames, *Annu. Rev. Contr.* **55**, 142 (2023).
61. M. H. Li, Y. Wen, X. Gao, J. Si, H. Huang, *IEEE Trans. Robot.* **38**, 407 (2022).

ACKNOWLEDGMENTS

K. G. acknowledges support of the US National Science Foundation, via the National Robotics Initiative (NRI), as part of the collaborative research project "Software Framework for Research in Semi-Autonomous Teleoperation" between The Johns Hopkins University (IIS 1637789), Worcester Polytechnic Institute (IIS 1637759), and the University of Washington (IIS 1637444). A.K. acknowledges support from the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the NIH under award R01EB020610 for "Next Generation of Surgical Imaging and Robotics for Supervised-Autonomous Soft Tissue Surgery" and the NSF CAREER award 2144348 for "Advancing Autonomy for Soft Tissue Robotic Surgery and Interventions."

10.1126/science.adj3312

On 15 March 2016, professional Go player Lee Sedol reviews the match after being defeated by AlphaGo.

PHOTO: LEE JIN-MAN/AP PHOTO

PERSPECTIVE

Improving artificial intelligence with games

Games continue to drive progress in the development of artificial intelligence

By **Peter R. Wurman[1], Peter Stone[1,2], Michael Spranger[3]**

For thousands of years, humans have played board games as a pleasurable way to pass time and match wits. Games such as Senet were so prized as to be buried with other valuables for the afterlife in Egyptian tombs ~3500 BCE (*1*). Skill in strategic board games such as chess and Go came to represent a pinnacle of human intelligence, and thus it is no surprise that such board games have been viewed as yardsticks by which to measure machine intelligence (*2*). In recent years, computers have overtaken human skill in a wide array of strategy games. Although it may seem that the recent dominance of artificial intelligence (AI) in classic strategy games foreshadows the end of what games can teach us about AI, researchers are just beginning to tackle the substantially more complex and varied challenges posed by modern video games.

Games provide safe opportunities to isolate and practice many skills that are transferable to the real world. As such, they are valuable training grounds not only for people but also for intelligent machines. Strategic reasoning and problem solving can be learned from traditional games such as tic-tac-toe, checkers, chess, and Go. Probabilistic reasoning can be learned from games such as backgammon, bridge, and poker. Social games such as Hanabi and Diplomacy teach cooperation and negotiation. Furthermore, video games substantially open up the space of skills that can be practiced, including running businesses (lemonade stands or railroads), managing growing simulated cities, or racing virtual cars.

Games such as checkers, chess, and Go represent the simplest class of games in that players have perfect information (all players see all of the information); there are no hidden or stochastic elements. In theory, the game can be "solved" by computing the optimal play for both players, which would guarantee a win or draw. Checkers (*3*), Connect Four, Gomoku, and nine men's morris are examples of mathematically solved games. Chess and Go remain unsolved because of their number of possible board states.

A weaker, but still impressive, demonstration of skill is to beat the best human players in the world, as when Deep Blue defeated reigning world chess champion Gary Kasparov in 1997 (*4*). In 2016, AlphaGo demonstrated similar superhuman performance in the game of Go (*5*). In tackling Go, a new stochastic search algorithm called Monte Carlo tree search (MCTS) was developed to handle large state spaces, which has had benefits for a variety of real-world optimization problems such as multirobot path planning and procedural content generation.

In large games, MCTS alone cannot sample enough possible futures to determine the value of early and middle game board positions accurately. Instead, the search must be complemented with good estimates of the long-term value of intermediate game states. Notably, superhuman performance in Go was not achieved until neural networks were employed to estimate the value of intermediate board positions, which made MCTS more accurate and efficient.

Although a computer defeated a sitting world backgammon champion in 1979 (*6*), decades before computers won at chess or Go, backgammon represents a slightly more challenging class of game because of the addition of random chance. In games of chance, the value of a node in the search tree must represent the expected value of that game state, that is, a weighted sum over all possible ways in which the stochastic future of the game might unfold. The backgammon program's performance was a high-profile demonstration of the power of using self-play and reinforcement learning (RL) to train a computer program.

Games such as poker and Stratego add the complexity that the player has imperfect information. Although a poker player knows their own cards, they do not know the opponents' cards or any shared cards that have not been exposed. Furthermore, in poker competitions, the goal is to win the most money over the long term, not just to win individual hands. Thus, an AI system must reason about not only the opponents' current cards, but

[1]Sony Research, New York, NY, USA. [2]The University of Texas at Austin, Austin, TX, USA. [3]Sony Research, Tokyo, Japan. Email: peter.wurman@sony.com; pstone@cs.utexas.edu; michael.spranger@sony.com

also their long-term betting strategies. In game theory terms, it is an example of a repeated game, where the interactions in one round affect the beliefs and strategies of the players in future rounds. Superhuman performance has been demonstrated in imperfect information games such as Stratego and Texas hold'em no-limit poker (*7*). However, interestingly, bridge remains a rare example of a game that AI has not mastered, possibly owing to the complex nature of the inter- and intrateam communication required to play well with a partner.

In an attempt to study more generally intelligent "agents," annual competitions were held between 2005 and 2016 to motivate research in game playing (*8*). The agents were presented with a logical definition of a new game in a formal game-description language and had to reason about not only how to play, but also how to represent the game in memory so that it can be searched and evaluated. More recently, MuZero (*9*), a single, unified system that could learn to play Atari, chess, Go, or shogi, was presented. It was followed soon after by Player of Games, which expanded the set to include imperfect-information games such as poker.

Although the accomplishments of designing computer programs that can defeat the best players in the world is impressive, strategy games require a particular kind of intelligence. Specifically, these successful programs use some form of guided deep search and state-value estimation. Compared to humans, computers are much faster and more thorough at search, and recently became better at state-value estimation by using self-play to generate vast amounts of expert-level data that is fed into neural networks that serve as function approximators.

By contrast, video games pose new types of challenges, many of which drive the development of systems that require more "common-sense" intelligence (*10*). Video games are a promising application area for work in general intelligence because they are complex simulations of an imaginary world, complete with (imaginary) physics and usually populated with a variety of other players and nonplayer characters. They entail many of the challenges associated with building general-purpose robots without the real-world complexities and costs. Indeed, developing agents to play video games allows researchers to discover and start to address undesirable emergent behaviors in relatively low-stakes environments, rather than in business-critical applications.

To date, only a few AI systems have been developed for these domains. In the late 1990s, games such as *Quake* and *Doom* were used to explore the power of rule-based AI systems. Things changed in 2015 when the confluence of progress in image processing through neural networks and RL enabled an agent to play some classic Atari games (*11*). Soon after, the suite of Atari games became a benchmark domain for AI researchers to compare RL algorithms. The Atari suite of games also exposed challenging situations, such as sparse rewards, that forced researchers to continue to expand the capabilities of the techniques. These advances enabled researchers to tackle more modern and difficult video games, including the real-time strategy games *StarCraft II* (*12*) and *DotA 2* (*13*), and the highly realistic racing game *Gran Turismo* (GT) (*14*). All three of these projects culminated with evaluations against top e-sports gamers in which they demonstrated grand-master or superhuman performance. In all of these successes, machine learning allowed researchers to achieve performance levels that they could not reach with hand-coded rules.

Although much has been learned about building AIs for these particular games, the challenges posed by open-world games such as *Minecraft*, *Horizon Forbidden West*, or *The Legend of Zelda: Tears of the Kingdom* have yet to be conquered. These games require AI systems to learn diverse skills and apply them across enormous maps with varied terrain, enemies, and other challenges. They involve elements of combat, travel, economics, and creative puzzle solving. They have complex storylines, communicated in natural language, that unfold throughout the gameplay and require the player to leverage acquired skills and knowledge to make progress. The techniques needed to build agents for small parts of these games are available, and recent progress in generative AI may help with some of these challenges. However, the task of building a general-purpose agent architecture that can learn—especially in a human time scale—to play these sorts of games remains a grand challenge.

Using video games as platforms for research in AI directly benefits the USD$200 billion global video game industry and, eventually, could benefit development of the metaverse. Producing agents that are capable of playing modern video games unlocks the possibilities of much more capable, human-like nonplayer characters within the games themselves. For example, anecdotal evidence from online forums regarding the recent, time-limited deployment of the racing agent (GT Sophy) in *Gran Turismo 7* suggests that players felt that racing against GT Sophy provided an experience that was much more like racing other skilled humans than the game's default, hand-coded AI. As the technology progresses, the gaming community could benefit from more-intelligent agents by using them as opponents and teammates in online games, as in-game tutors, as automated testers to find bugs in games, or by designers to ensure game balance.

In the physical world, games often take the form of sports. Research on robots that can play sports, such as the ongoing RoboCup (*15*) robotic soccer competitions and the autonomous race cars being developed in the Indy Autonomous Challenge, simultaneously spur progress on the frontiers of AI and robotics. In both examples, much of the research occurs in simulations because the cost and complexity of keeping real robots running preclude them from being used as the sole experimental platform. Bridging the sim-to-real gap, a concept that describes the challenges of moving a policy from simulation to real robots, has become a topic of research on these projects. Fortunately, owing in large part to decades of efforts to improve the graphical and physical realism of video games, the fidelity of robot simulations has continued to improve, thereby reducing sim-to-real gaps.

Making progress on AI within video games will also be a substantial step toward much more capable and flexible AI systems that operate in the physical world. In contrast to special-purpose robots, general-purpose robots need the ability to continually adapt to their environments and quickly learn new tasks. Modern, open-world video games allow AI researchers to focus on fundamental AI topics such as multitask learning, life-long learning, common-sense reasoning, human-AI collaboration, and explainability, while leaving the challenges raised by real-world sensors and actuators to roboticists. As robotic simulators become even more realistic, the sim-to-real gap will continue to shrink, making it more likely that the techniques that work in games and simulations will be transferable to the real world. ■

REFERENCES AND NOTES

1. W. Crist *et al.*, *Ancient Egyptians at Play: Board Games Across Borders* (Bloomsbury, 2016).
2. N. Wiener, *Cybernetics* (Wiley, 1948).
3. J. Schaeffer *et al.*, *Science* **317**, 1518 (2007).
4. M. Campbell, A. J. Hoane Jr., F. Hsu, *Artif. Intell.* **134**, 57 (2002).
5. D. Silver *et al.*, *Nature* **529**, 484 (2016).
6. G. Tesauro, *Neural Comput.* **6**, 215 (1994).
7. N. Brown, T. Sandholm, *Science* **359**, 418 (2018).
8. M. Genesereth, Y. Björnsson, *AI Mag.* **34**, 107 (2013).
9. J. Schrittwieser *et al.*, *Nature* **588**, 604 (2020).
10. J. Laird *et al.*, *AI Mag.* **22**, 15 (2001).
11. V. Mnih *et al.*, *Nature* **518**, 529 (2015).
12. O. Vinyals *et al.*, *Nature* **575**, 350 (2019).
13. M. Hutson, *Science* **361**, 632 (2018).
14. P. R. Wurman *et al.*, *Nature* **602**, 223 (2022).
15. H. Kitano *et al.*, *Proceedings. of the First International Conference on Autonomous Agents* (1997), pp. 340–347.

ACKNOWLEDGMENTS

All authors are coauthors on patents and receive stock options as employees of Sony. The employment of P.S. at Sony has been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

10.1126/science.adh8135

PERSPECTIVE - HYPOTHESIS

Implications of predicting race variables from medical images

AI-predicted race variables pose risks and opportunities for studying health disparities

By **James Zou[1], Judy Wawira Gichoya[2], Daniel E. Ho[3], Ziad Obermeyer[4]**

There are now more than 500 US Food and Drug Administration (FDA)–approved medical artificial intelligence (AI) devices, and AI is used in diverse medical tasks such as assessing the risk of heart failure and diagnosing cancer from images (*1*). Beyond predicting standard diagnoses, AI models can infer a substantial number of patient features from medical images in ways that humans cannot. For example, several studies have demonstrated that some AI models can infer race variables (crude simplistic categories) directly from medical images such as chest x-rays and cardiac ultrasounds, even though there are no known human-readable race correlates in these images (*2*–*4*). This has raised concerns about the possibility of AI systems to discriminate. At the same time, AI predictions of demographic attributes, including race variables, also have the potential to help assess and monitor health care disparities and generate new insights into risk factors (*5*).

Race is a complex and problematic social construct that does not have a biological basis (*6*). The concept of race varies across time and geography and according to social factors. Demographic-prediction AI models can impute (or infer) race variables that are captured in electronic health records. These are simplistic and dated reporting categories, such as white, Black, and Asian. Despite such limitations, measures of health care equity and algorithmic fairness often depend on stratifying outcomes by race variables.

How AI predicts race variables is puzzling. It is likely that other demographic attributes, comorbidities, or imaging artifacts are at play. For example, AI can predict patients' age and sex from cardiac ultrasound videos because heartbeat patterns vary with age and sex (*3*). Therefore, if patients of one race skew older or younger in the training dataset, then the model can use this correlation to predict race. Measured confounders (variables that may cause spurious association) that could lead to correlations can be adjusted for directly. But in a study of chest x-rays, measured potential confounders were balanced across the dataset, and a deep learning model was still able to predict race variables from the images, achieving an area under the receiver operating characteristic curve (AUROC) above 0.9 (AUROC is a common metric of prediction accuracy where 1 is perfect) (*2*). This suggests that unmeasured confounders also play a role. Moreover, illnesses can be unequally distributed across racial groups; if those illnesses have imaging manifestations, algorithms could also use them to predict race.

"...the ability of AI to predict race variables from medical images could be useful for monitoring health care disparity..."

Could the ability of AI to predict race variables from medical images, whatever the mechanism, exacerbate health care disparities? Although diagnostic AI is not trained to directly read out race variables, it is possible that race or its correlates could be a shortcut that is used by AI algorithms to make disease diagnoses. There are some settings where this is possible and potentially harmful. For example, suppose the patients of one race are more likely to have their images taken by type A machines and, because of differences in where the machines are deployed, most of the positive cases in the training data are collected from type B machines. AI can learn to use machine type to predict race variables and disease status (it is less likely to predict disease for images taken from machine A). In deployment, the AI would then underdiagnose patients of a particular race, leading to care disparity. This raises serious questions of algorithmic discrimination, especially if the AI uses race proxy in its assessment and clinicians are not aware of this.

Therefore, it is critical to evaluate the reliability of AI in truly independent and representative datasets in which race variables are available—for example, from patient self-reporting or linked administrative data. Ideally, this dataset would come from multiple settings or health systems to avoid overfitting. Then, equity can be measured by disaggregating the model's performance by race variables to ensure that it does not have excess errors or differential performance between groups.

The ability of AI to pick up hidden phenotypes, such as race variables, from medical data could also be a resource for equity. AI can generate valuable information for auditing datasets to detect biases (*7*). For example, if an AI model can accurately predict attributes that may not be medically relevant (such as race, or the machine type that took the image), this suggests possible confounders in the training dataset. Maybe there are spurious correlations involving race variables in the data, or parts of the hospital have different imaging setups. In such cases, these confounders should be explicitly accounted for in subsequent analyses. Unexpected predictions by AI force these data warts into the open and could improve transparency.

In many settings, race variables are not available, but laws in numerous countries require assessment of disparities. How can it be ensured that a model, treatment, or policy benefits a diverse population if the relevant demographic attributes are missing? Predicted race variables from an AI algorithm could be a useful, albeit imprecise, proxy. For example, if researchers develop a method to diagnose Long Covid from chest x-ray images, but they do not have race metadata for many images in their dataset, a demographic-prediction AI model could impute race variables from their images. The model's performance when stratified by race variables can then be assessed to ensure that it works well across diverse populations. There is precedent for this imputation approach: The US Centers for Medicare and Medicaid Services use race inferred from surname and location to monitor health care disparity (*8*), and the Consumer Financial Protection Bureau uses a similar approach to monitor equity in lending (*9*). The Internal Revenue Service is statutorily prohibited from using Census demographic data to assess disparities in tax administration, and the US Treasury Department has in turn relied on imputed race variables to

[1]Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. [2]Department of Radiology, Emory University, Atlanta, GA, USA. [3]Stanford Law School, Stanford University, Stanford, CA, USA. [4]School of Public Health, University of California, Berkeley, Berkeley, CA, USA. Email: jamesz@stanford.edu

uncover important disparities in tax expenditures and audits (*10*).

The usage of imputed race variables in health care should focus on disparity assessment and algorithmic auditing instead of medical decision-making. The accuracy of imputations should be carefully validated in each new application. When the imputed race variables are broad categories, downstream analysis may miss disparities within subgroups of the population. Furthermore, in applications involving interpretations of DNA sequencing, other population labels such as genetic ancestry are more meaningful categories for assessment than race variables (*11*).

AI could help to provide more nuanced representations of race than the standard reporting categories, which are often very coarse. AI produces continuous scores or probabilities, not categories. This mirrors work in human genetics showing that it is more accurate to model populations as continuous spectrums of variation in ancestry rather than modeling according to a few discrete race labels (*12*). AI models, such as neural networks, also learn to represent individuals as points in a continuous space (often called embedding) (*13*). Instead of treating, for example, "Asian" as a single discrete category, an AI model that assesses chest x-rays maps images from Asian patients to a continuous geometric space before making its diagnosis prediction. Variation in this space could reflect ancestry subgroups within Asians as well as differences in health care access, comorbidities, and other unmeasured factors. More research is needed to characterize the implications for health care from differences between patient embedding.

AI represents patients on a continuous spectrum rather than as discrete racial categories. This makes it easier to evaluate and ensure that the model makes accurate predictions across fine-grained subpopulations. Instead of measuring how well the AI model performs on Asians as one group, which can mask limitations on less populous subgroups, the model's performance can be assessed on different clusters or spectra of patients in the embedding space. Suppose the model makes more mistakes than expected in a particular data cluster; further investigation of what subpopulation is represented in this error cluster can be undertaken, and the model could be retrained to improve on this subgroup. In response to advocacy by civil rights groups, the White House recently proposed updating the 1997 coarse race reporting standard to foster data disaggregation according to reporting outcomes by subgroups (for example, Vietnamese American versus Asian American). It may take years for that process to materialize in health data, and AI imputations can potentially help to uncover health care disparities between granular subgroups in the meantime (*14*).

Although race variables are not a generally meaningful category in medicine, the ability of AI to predict race variables from medical images could be useful for monitoring health care disparity and ensuring that algorithms work well across diverse populations. Where feasible, researchers can try to predict race and other demographic variables from their data as a mechanism to thoroughly audit their dataset and model. If race variables and other attributes can be accurately predicted, this could reveal potential data or model biases that should be addressed. Understanding what features AI mechanistically uses to predict race variables is important, and AI interpretation methods such as concept activation, saliency map, and counterfactual explanations could help. Overall, being transparent and deliberate about race variable predictions should reduce the risk of algorithmic discrimination and could generate valuable information about the dataset, the model, and health care disparities. ■

REFERENCES AND NOTES

1. E. Wu *et al.*, *Nat. Med.* **27**, 582 (2021).
2. J. W. Gichoya *et al.*, *Lancet Digit Health*. **4**, e406 (2022).
3. G. Duffy *et al.*, *NPJ Digit. Med.* **5**, 188 (2022).
4. J. Adleberg *et al.*, *J. Am. Coll. Radiol.* **19**, 1151 (2022).
5. R. Poplin *et al.*, *Nat. Biomed. Eng.* **2**, 158 (2018).
6. A. Deyrup, J. L. Graves Jr., *N. Engl. J. Med.* **386**, 501 (2022).
7. W. Liang *et al.*, *Nat. Mach. Intell.* **4**, 669 (2022).
8. A. Fremont, J. S. Weissman, E. Hoch, M. N. Elliott, *Rand Health Q.* **6**, 16 (2016).
9. Consumer Financial Protection Bureau, Using publicly available information to proxy for unidentified race and ethnicity (2014); https://www.consumerfinance.gov/data-research/research-reports/using-publicly-available-information-to-proxy-for-unidentified-race-and-ethnicity.
10. J.-A. Cronin, P. DeFilippes, R. Fisher, Tax Expenditures by Race and Hispanic Ethnicity: An application of the U.S. Treasury Department's race and hispanic ethnicity imputation, Department of Treasury Office of Tax Analysis working paper (2023).
11. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research, Board on Health Sciences Policy, Committee on Population, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine, *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* (National Academies, 2023).
12. M. Yudell, D. Roberts, R. DeSalle, S. Tishkoff, *Science* **351**, 564 (2016).
13. Y. LeCun, Y. Bengio, G. Hinton, *Nature*. **521**, 436 (2015).
14. L. Cheng, I. Gallegos, D. Ouyang, J. Goldin, D. E. Ho, *Proceeding of the ACM Conference on Fairness, Accountability, and Transparency* 10.1145/3593013.3594034 (2023).

10.1126/science.adh4260

PERSPECTIVE

Mitigating bias in AI at the point of care

Promoting equity in AI in health care requires addressing biases at clinical implementation

By **Matthew DeCamp**[1] *and* **Charlotta Lindvall**[2,3,4]

Artificial intelligence (AI) shows promise for improving basic and translational science, medicine, and public health, but its success is not guaranteed. Numerous examples have arisen of racial, ethnic, gender, disability, and other biases in AI applications to health care. In ethics, bias generally refers to any systematic, unfair favoring of people in terms of how they are treated or the outcomes they experience. Consensus has emerged among scientists, ethicists, and policy-makers that minimizing bias is a shared responsibility among all involved in AI development. For example, ensuring equity by eliminating bias in AI is a core principle of the World Health Organization for governing AI (*1*). But ensuring equity will require more than unbiased data and algorithms. It will also require reducing biases in how clinicians and patients use AI-based algorithms—a potentially more challenging task than reducing biases in algorithms themselves.

Examination of bias in AI has tended toward removing bias from datasets, analyses, or AI development teams. For example, because of unequal recruitment and enrollment, oncology datasets demonstrate racial, ethnic, and global geographic biases (*2*). In another example, developers assumed that health care costs were a proxy for health care needs, but then learned that Black Americans often receive less medical care even when they have greater

[1]Center for Bioethics and Humanities, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [2]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Medicine, Harvard Medical School, Boston, MA, USA. Email: matthew.decamp@cuanschutz.edu

uncover important disparities in tax expenditures and audits (*10*).

The usage of imputed race variables in health care should focus on disparity assessment and algorithmic auditing instead of medical decision-making. The accuracy of imputations should be carefully validated in each new application. When the imputed race variables are broad categories, downstream analysis may miss disparities within subgroups of the population. Furthermore, in applications involving interpretations of DNA sequencing, other population labels such as genetic ancestry are more meaningful categories for assessment than race variables (*11*).

AI could help to provide more nuanced representations of race than the standard reporting categories, which are often very coarse. AI produces continuous scores or probabilities, not categories. This mirrors work in human genetics showing that it is more accurate to model populations as continuous spectrums of variation in ancestry rather than modeling according to a few discrete race labels (*12*). AI models, such as neural networks, also learn to represent individuals as points in a continuous space (often called embedding) (*13*). Instead of treating, for example, "Asian" as a single discrete category, an AI model that assesses chest x-rays maps images from Asian patients to a continuous geometric space before making its diagnosis prediction. Variation in this space could reflect ancestry subgroups within Asians as well as differences in health care access, comorbidities, and other unmeasured factors. More research is needed to characterize the implications for health care from differences between patient embedding.

AI represents patients on a continuous spectrum rather than as discrete racial categories. This makes it easier to evaluate and ensure that the model makes accurate predictions across fine-grained subpopulations. Instead of measuring how well the AI model performs on Asians as one group, which can mask limitations on less populous subgroups, the model's performance can be assessed on different clusters or spectra of patients in the embedding space. Suppose the model makes more mistakes than expected in a particular data cluster; further investigation of what subpopulation is represented in this error cluster can be undertaken, and the model could be retrained to improve on this subgroup. In response to advocacy by civil rights groups, the White House recently proposed updating the 1997 coarse race reporting standard to foster data disaggregation according to reporting outcomes by subgroups (for example, Vietnamese American versus Asian American). It may take years for that process to materialize in health data, and AI imputations can potentially help to uncover health care disparities between granular subgroups in the meantime (*14*).

Although race variables are not a generally meaningful category in medicine, the ability of AI to predict race variables from medical images could be useful for monitoring health care disparity and ensuring that algorithms work well across diverse populations. Where feasible, researchers can try to predict race and other demographic variables from their data as a mechanism to thoroughly audit their dataset and model. If race variables and other attributes can be accurately predicted, this could reveal potential data or model biases that should be addressed. Understanding what features AI mechanistically uses to predict race variables is important, and AI interpretation methods such as concept activation, saliency map, and counterfactual explanations could help. Overall, being transparent and deliberate about race variable predictions should reduce the risk of algorithmic discrimination and could generate valuable information about the dataset, the model, and health care disparities. ■

REFERENCES AND NOTES

1. E. Wu *et al.*, *Nat. Med.* **27**, 582 (2021).
2. J. W. Gichoya *et al.*, *Lancet Digit Health*. **4**, e406 (2022).
3. G. Duffy *et al.*, *NPJ Digit. Med.* **5**, 188 (2022).
4. J. Adleberg *et al.*, *J. Am. Coll. Radiol.* **19**, 1151 (2022).
5. R. Poplin *et al.*, *Nat. Biomed. Eng.* **2**, 158 (2018).
6. A. Deyrup, J. L. Graves Jr., *N. Engl. J. Med.* **386**, 501 (2022).
7. W. Liang *et al.*, *Nat. Mach. Intell.* **4**, 669 (2022).
8. A. Fremont, J. S. Weissman, E. Hoch, M. N. Elliott, *Rand Health Q.* **6**, 16 (2016).
9. Consumer Financial Protection Bureau, Using publicly available information to proxy for unidentified race and ethnicity (2014); https://www.consumerfinance.gov/data-research/research-reports/using-publicly-available-information-to-proxy-for-unidentified-race-and-ethnicity.
10. J.-A. Cronin, P. DeFilippes, R. Fisher, Tax Expenditures by Race and Hispanic Ethnicity: An application of the U.S. Treasury Department's race and hispanic ethnicity imputation, Department of Treasury Office of Tax Analysis working paper (2023).
11. Committee on the Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research, Board on Health Sciences Policy, Committee on Population, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, National Academies of Sciences, Engineering, and Medicine, *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field* (National Academies, 2023).
12. M. Yudell, D. Roberts, R. DeSalle, S. Tishkoff, *Science* **351**, 564 (2016).
13. Y. LeCun, Y. Bengio, G. Hinton, *Nature*. **521**, 436 (2015).
14. L. Cheng, I. Gallegos, D. Ouyang, J. Goldin, D. E. Ho, *Proceeding of the ACM Conference on Fairness, Accountability, and Transparency* 10.1145/3593013.3594034 (2023).

10.1126/science.adh4260

PERSPECTIVE

Mitigating bias in AI at the point of care

Promoting equity in AI in health care requires addressing biases at clinical implementation

By **Matthew DeCamp**[1] *and* **Charlotta Lindvall**[2,3,4]

Artificial intelligence (AI) shows promise for improving basic and translational science, medicine, and public health, but its success is not guaranteed. Numerous examples have arisen of racial, ethnic, gender, disability, and other biases in AI applications to health care. In ethics, bias generally refers to any systematic, unfair favoring of people in terms of how they are treated or the outcomes they experience. Consensus has emerged among scientists, ethicists, and policy-makers that minimizing bias is a shared responsibility among all involved in AI development. For example, ensuring equity by eliminating bias in AI is a core principle of the World Health Organization for governing AI (*1*). But ensuring equity will require more than unbiased data and algorithms. It will also require reducing biases in how clinicians and patients use AI-based algorithms—a potentially more challenging task than reducing biases in algorithms themselves.

Examination of bias in AI has tended toward removing bias from datasets, analyses, or AI development teams. For example, because of unequal recruitment and enrollment, oncology datasets demonstrate racial, ethnic, and global geographic biases (*2*). In another example, developers assumed that health care costs were a proxy for health care needs, but then learned that Black Americans often receive less medical care even when they have greater

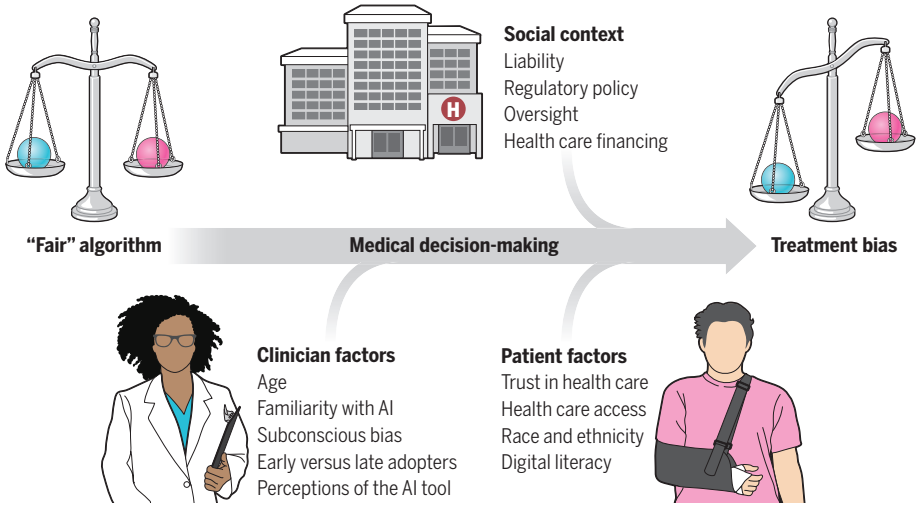
[1]Center for Bioethics and Humanities, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [2]Department of Psychosocial Oncology and Palliative Care, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Department of Informatics and Analytics, Dana-Farber Cancer Institute, Boston, MA, USA. [4]Department of Medicine, Harvard Medical School, Boston, MA, USA. Email: matthew.decamp@cuanschutz.edu

need (*3*); the resulting algorithm would have cemented this structural racism into place. Moreover, encouraging diversity and inclusivity in researchers is another way to manage AI bias (*4*).

Alone these efforts are insufficient. Mitigating bias when implementing AI suffers from a more challenging problem: emergent biases, i.e., biases that emerge not merely because of biased datasets or algorithms, but because of factors involved in real-world implementation (*5*). The benefits or harms of an algorithm, even one agreed upon as "fair," are likely to be experienced unequally by patients. These harms can be seen as latent biases (i.e., biases waiting to happen) and the factors involved as latent conditions [i.e., characteristics of the environment that contribute to bias (*6*)]. Clinician-, patient-, and social-level factors can interact to create biases (see the figure).

How a "fair" algorithm can result in biased outcomes

A clinician and patient interact with artificial intelligence (AI)–based decision support that provides information about, for example, the likelihood of a diagnosis, utility of a treatment, or a prognosis. Even if the algorithm is unbiased, clinician-, patient-, and social-level factors can influence how the recommendations are interpreted and implemented, which can result in latent treatment bias.



GRAPHIC: K. HOLOSKI/*SCIENCE*

Clinician-level factors will play complex roles in determining whether and how clinicians choose to use AI-derived information. Most physicians do not blindly adopt and follow clinical decision support systems (CDSSs) or recommendations based on algorithms. However, the impacts of various factors are not always obvious.

Although not specific to AI, in some prior studies, older physicians are less likely to follow algorithm-based recommendations, whereas in other studies, younger physicians override decision support more (*7*). Thus, clinical experience or perceived expertise alone may not predict willingness to follow AI recommendations. Still, early adopters of technology will be more likely to incorporate AI-based recommendations compared to late adopters, meaning that patients cared for by early adopters will be more likely to experience benefits or harms of AI sooner. Clinicians familiar with AI may be more (or less) likely to incorporate AI. Age, experience, and other factors will thus influence whether or not a clinician follows an AI-based recommendation, though perhaps not in a consistent pattern.

Moreover, that a model is "AI" could influence clinicians. Traditional CDSSs are based on expert knowledge and guidelines, whereas AI-based models rely primarily on statistical associations. Statistical associations can be built upon variables or their interactions that clinicians might not know to include in their decision-making. AI algorithms (e.g., deep neural networks) often lack transparency or interpretability regarding how data are transformed into model outputs, sometimes referred to as "black box" algorithms. Because medicine places value on expert and experiential knowledge, statistical or "black box" knowledge could be discounted in decisions. How clinicians perceive AI-based tools—in terms of how a recommendation is presented, transparency surrounding the motivation behind the AI and its data sources, and its explainability—could influence clinicians' willingness to use AI. As a result, differential use and uptake of AI should be expected among clinicians. This means that patients under individual clinicians' care will experience the benefits and risks of AI in systematically different ways—an issue of potential bias.

Complicating matters further, patient-level factors will affect how patients experience AI. For one, because of socioeconomic status, race, ethnicity, geography, and other factors, inequities in access to clinicians, hospitals, and other services are a problem for all health systems worldwide. This includes unequal access to AI-based technology. Additionally, some patients may trust technology more than others, and some may have less technology literacy. Patients who distrust a clinician or health system may be less likely to trust that clinician's or health system's use of AI, and patient distrust is not randomly distributed: It associates with race, ethnicity, education, and other factors. If given the choice, such patients may simply opt out of AI in their care (*8*).

The severity of patients' illnesses or symptom type may affect clinicians' willingness to follow computer-based recommendations. For example, for a patient in the United States who identifies as Black and presents with a symptom of pain, data suggest that clinicians may be more likely to override decision support based on a patient's symptom [e.g., certain types of pain (*7*)] and also to interpret reports of pain differently based on race (*9*). Therefore, differential following of AI-recommended treatment plans according to patient race could differentially harm or help Black patients.

A third source of factors that contribute to emergent biases comes from the social and policy context. Health care financing mechanisms vary widely, and the affordability and accessibility of AI and what it recommends will influence its effects. Because medical liability is defined on the basis of current standards of care that do not yet involve AI, the safest way for clinicians to use AI will be as confirmation for the existing standard of care (*10*). Early uses of AI may thus suffer from status quo bias, hindering the potential of AI to result in innovation or changes to care plans for patients.

Clinicians may either rely more on the technology or ignore it completely in time-pressured clinical environments. This automation bias (over- or under-reliance on the AI) may exacerbate disparities for under-resourced settings where time and financial pressures are more intense. In some settings, clinicians may be more willing to trust and use AI that is developed or endorsed by their own hospital or health system, compared to commercially developed AI.

As latent biases emerge from clinical implementation, they will feed back into learning models, potentially negatively affecting

the performance of the model itself (*11*). Increasing clinician awareness of AI's biases is critical, but this desire may be paradoxical: Knowing about biases in AI may result in less willingness to use AI-based recommendations for patients that a clinician judges "different" from others. Assuming models are biased in terms of race or ethnicity, for example, could result in clinicians systematically overriding a model's recommendation for that group of patients.

Several strategies exist to identify and address latent biases. One strategy could involve providing clinicians with model-specific, individual-level performance feedback regarding whether they tend to outperform or underperform it, or if they are systematically following or overriding a model only for certain patient groups. Individualized feedback has the potential to improve clinician performance (*12*). However, a challenge for assessing bias is that clinicians may not see sufficient numbers of patients in different groups to allow rigorous, stratified comparisons.

Patients should be informed about the use of AI in their clinical care as a matter of respect. This includes general messaging about the use of predictive algorithms, chatbots, and other AI-based technologies, and specific notification when new AI-based technologies are used in their individual care. Doing so may improve awareness of AI, motivate conversations with clinicians, and support greater transparency around AI use.

Exactly how much to disclose, and in what format, are unanswered questions that require additional research. There is a need to avoid AI exceptionalism—the idea that AI is riskier or requires greater protection, just because it is AI—and presently patients want to know more, not less (*8*). That other decisions relying on algorithms, such as clinical risk calculators or computer-aided radiographic or electrocardiogram interpretation, may not be routinely shared with patients is not an argument in favor of secrecy.

Bias has not been a major aspect of drug and device regulations, which focus on overall safety and efficacy. Recent US proposals could extend legal liability to physicians and hospitals, meaning they could be required to provide compensatory damages to patients or be subject to penalties for use of biased clinical algorithms; these could be applied to AI algorithms (*13*). However, the complexity of AI algorithms and persistent ethical disagreement over when differential performance by race or ethnicity equals true bias complicate liability proposals. Drug and device regulatory agencies might consider making evaluations of bias mandatory for approval (*14*). A first step could be requiring evaluations of differential performance and bias under different real-world assumptions in approval processes and other forums, such as in journal reporting of AI research.

In addition, the gaze of AI should be turned on itself. This requires proactive, intentional development of AI tools to identify biases in AI and in its clinical implementation (*15*). AI may contribute to the emergence of biases, but it also has the potential to detect biases and hence facilitate new ways of overcoming them. Open-source tools, such as AI Fairness 360, FairML, and others, show promise in helping researchers assess fairness in their machine learning data and algorithms. These tools can assess biases in datasets, predictive outputs, and even the different techniques that can be used to mitigate bias according to different metrics of fairness. Their application to health care data and algorithms deserves rigorous scientific examination.

Implementation research is urgently needed to better understand the role of different contextual factors and latent conditions in allowing biases to emerge. Exactly which patients may experience bias under which circumstances requires ongoing rigorous study. In AI, biased data and biased algorithms result in biased outcomes for patients, but so do unbiased data and algorithms when they enter a biased world. All patients deserve to benefit from both fair algorithms and fair implementation. ■

REFERENCES AND NOTES

1. World Health Organization, *Ethics and Governance of Artificial Intelligence for Health* (World Health Organization, 2021).
2. S. Guerrero *et al.*, *Sci. Rep.* **8**, 13978 (2018).
3. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **366**, 447 (2019).
4. S. M. West *et al.*, Discriminating Systems: Gender, Race and Power in AI (AI Now Institute, 2019).
5. A. J. London, *Cell Rep. Med.* **3**, 100622 (2022).
6. C. M. Lowe, *Qual. Saf. Health Care* **15** (suppl. 1), i72 (2006).
7. J. Yoo *et al.*, *JMIR Med. Inform.* **8**, e23351 (2020).
8. A. Tyson *et al.*, "60% of Americans Would Be Uncomfortable with Provider Relying on AI in Their Own Health Care" (Pew Research Center, 2023).
9. D. P. Ly, *JAMA Health Forum* **2**, e212333 (2021).
10. W. N. Price 2nd, S. Gerke, I. G. Cohen, *JAMA* **322**, 1765 (2019).
11. A. Bracic, S. L. Callier, W. N. Price 2nd, *Science* **377**, 1158 (2022).
12. N. Ivers *et al.*, *Cochrane Database Syst. Rev.* **6**, CD000259 (2012).
13. K. E. Goodman, D. J. Morgan, D. E. Hoffmann, *JAMA* **329**, 285 (2023).
14. K. Dortche *et al.*, *J. Sci. Policy Gov.* 10.38126/JSPG210302 (2023).
15. R. B. Parikh, S. Teeple, A. S. Navathe, *JAMA* **322**, 2377 (2019).

ACKNOWLEDGMENTS

The authors thank K. Tang for help with the illustration, M. Akerson for references, and M. Wynia and L. Hunter for comments on earlier drafts. M.D. receives funding from the NIH (grant R01NR019782) and Greenwall Foundation.

10.1126/science.adh2713

PERSPECTIVE

Using machine learning to decode animal communication

New methods promise transformative insights and conservation benefits

By **Christian Rutz[1], Michael Bronstein[2,3], Aza Raskin[4], Sonja C. Vernes[1,5], Katherine Zacarian[4], Damián E. Blasi[6]**

The past few years have seen a surge of interest in using machine learning (ML) methods for studying the behavior of nonhuman animals (hereafter "animals") (*1*). A topic that has attracted particular attention is the decoding of animal communication systems using deep learning and other approaches (*2*). Now is the time to tackle challenges concerning data availability, model validation, and research ethics, and to embrace opportunities for building collaborations across disciplines and initiatives.

Researchers must infer the meaning, or function, of animal signals through observation and experimentation (*3*). This is a challenging task, not least because animals use a wide range of communication modalities, including visual, acoustic, tactile, chemical, and electrical signals—often in conjunction, and beyond humans' perceptive capabilities. Observational work focuses on recording the signals of interest as well as detailed contextual information, including the identity, state, and behavior of both the senders and receivers of signals, their relationships and past interactions, and relevant environmental conditions. Some signal types may be produced only under certain circumstances, eliciting a specific behavioral response; a classic example is a vervet monkey (*Chlorocebus pygerythrus*) giving an alarm call when it spots a predator, which causes group members to

[1]School of Biology, University of St Andrews, St Andrews, Scotland, UK. [2]Department of Computer Science, University of Oxford, Oxford, UK. [3]Project CETI, New York, NY, USA. [4]Earth Species Project, Berkeley, CA, USA. [5]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. [6]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. Email: christian.rutz@st-andrews.ac.uk

the performance of the model itself (*11*). Increasing clinician awareness of AI's biases is critical, but this desire may be paradoxical: Knowing about biases in AI may result in less willingness to use AI-based recommendations for patients that a clinician judges "different" from others. Assuming models are biased in terms of race or ethnicity, for example, could result in clinicians systematically overriding a model's recommendation for that group of patients.

Several strategies exist to identify and address latent biases. One strategy could involve providing clinicians with model-specific, individual-level performance feedback regarding whether they tend to outperform or underperform it, or if they are systematically following or overriding a model only for certain patient groups. Individualized feedback has the potential to improve clinician performance (*12*). However, a challenge for assessing bias is that clinicians may not see sufficient numbers of patients in different groups to allow rigorous, stratified comparisons.

Patients should be informed about the use of AI in their clinical care as a matter of respect. This includes general messaging about the use of predictive algorithms, chatbots, and other AI-based technologies, and specific notification when new AI-based technologies are used in their individual care. Doing so may improve awareness of AI, motivate conversations with clinicians, and support greater transparency around AI use.

Exactly how much to disclose, and in what format, are unanswered questions that require additional research. There is a need to avoid AI exceptionalism—the idea that AI is riskier or requires greater protection, just because it is AI—and presently patients want to know more, not less (*8*). That other decisions relying on algorithms, such as clinical risk calculators or computer-aided radiographic or electrocardiogram interpretation, may not be routinely shared with patients is not an argument in favor of secrecy.

Bias has not been a major aspect of drug and device regulations, which focus on overall safety and efficacy. Recent US proposals could extend legal liability to physicians and hospitals, meaning they could be required to provide compensatory damages to patients or be subject to penalties for use of biased clinical algorithms; these could be applied to AI algorithms (*13*). However, the complexity of AI algorithms and persistent ethical disagreement over when differential performance by race or ethnicity equals true bias complicate liability proposals. Drug and device regulatory agencies might consider making evaluations of bias mandatory for approval (*14*). A first step could be requiring evaluations of differential performance and bias under different real-world assumptions in approval processes and other forums, such as in journal reporting of AI research.

In addition, the gaze of AI should be turned on itself. This requires proactive, intentional development of AI tools to identify biases in AI and in its clinical implementation (*15*). AI may contribute to the emergence of biases, but it also has the potential to detect biases and hence facilitate new ways of overcoming them. Open-source tools, such as AI Fairness 360, FairML, and others, show promise in helping researchers assess fairness in their machine learning data and algorithms. These tools can assess biases in datasets, predictive outputs, and even the different techniques that can be used to mitigate bias according to different metrics of fairness. Their application to health care data and algorithms deserves rigorous scientific examination.

Implementation research is urgently needed to better understand the role of different contextual factors and latent conditions in allowing biases to emerge. Exactly which patients may experience bias under which circumstances requires ongoing rigorous study. In AI, biased data and biased algorithms result in biased outcomes for patients, but so do unbiased data and algorithms when they enter a biased world. All patients deserve to benefit from both fair algorithms and fair implementation. ■

REFERENCES AND NOTES

1. World Health Organization, *Ethics and Governance of Artificial Intelligence for Health* (World Health Organization, 2021).
2. S. Guerrero *et al.*, *Sci. Rep.* **8**, 13978 (2018).
3. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **366**, 447 (2019).
4. S. M. West *et al.*, Discriminating Systems: Gender, Race and Power in AI (AI Now Institute, 2019).
5. A. J. London, *Cell Rep. Med.* **3**, 100622 (2022).
6. C. M. Lowe, *Qual. Saf. Health Care* **15** (suppl. 1), i72 (2006).
7. J. Yoo *et al.*, *JMIR Med. Inform.* **8**, e23351 (2020).
8. A. Tyson *et al.*, "60% of Americans Would Be Uncomfortable with Provider Relying on AI in Their Own Health Care" (Pew Research Center, 2023).
9. D. P. Ly, *JAMA Health Forum* **2**, e212333 (2021).
10. W. N. Price 2nd, S. Gerke, I. G. Cohen, *JAMA* **322**, 1765 (2019).
11. A. Bracic, S. L. Callier, W. N. Price 2nd, *Science* **377**, 1158 (2022).
12. N. Ivers *et al.*, *Cochrane Database Syst. Rev.* **6**, CD000259 (2012).
13. K. E. Goodman, D. J. Morgan, D. E. Hoffmann, *JAMA* **329**, 285 (2023).
14. K. Dortche *et al.*, *J. Sci. Policy Gov.* 10.38126/JSPG210302 (2023).
15. R. B. Parikh, S. Teeple, A. S. Navathe, *JAMA* **322**, 2377 (2019).

ACKNOWLEDGMENTS

The authors thank K. Tang for help with the illustration, M. Akerson for references, and M. Wynia and L. Hunter for comments on earlier drafts. M.D. receives funding from the NIH (grant R01NR019782) and Greenwall Foundation.

10.1126/science.adh2713

PERSPECTIVE

Using machine learning to decode animal communication

New methods promise transformative insights and conservation benefits

By **Christian Rutz[1], Michael Bronstein[2,3], Aza Raskin[4], Sonja C. Vernes[1,5], Katherine Zacarian[4], Damián E. Blasi[6]**

The past few years have seen a surge of interest in using machine learning (ML) methods for studying the behavior of nonhuman animals (hereafter "animals") (*1*). A topic that has attracted particular attention is the decoding of animal communication systems using deep learning and other approaches (*2*). Now is the time to tackle challenges concerning data availability, model validation, and research ethics, and to embrace opportunities for building collaborations across disciplines and initiatives.

Researchers must infer the meaning, or function, of animal signals through observation and experimentation (*3*). This is a challenging task, not least because animals use a wide range of communication modalities, including visual, acoustic, tactile, chemical, and electrical signals—often in conjunction, and beyond humans' perceptive capabilities. Observational work focuses on recording the signals of interest as well as detailed contextual information, including the identity, state, and behavior of both the senders and receivers of signals, their relationships and past interactions, and relevant environmental conditions. Some signal types may be produced only under certain circumstances, eliciting a specific behavioral response; a classic example is a vervet monkey (*Chlorocebus pygerythrus*) giving an alarm call when it spots a predator, which causes group members to

[1]School of Biology, University of St Andrews, St Andrews, Scotland, UK. [2]Department of Computer Science, University of Oxford, Oxford, UK. [3]Project CETI, New York, NY, USA. [4]Earth Species Project, Berkeley, CA, USA. [5]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. [6]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. Email: christian.rutz@st-andrews.ac.uk

Machine learning is helping to map and understand the vocal repertoire of tool-using Hawaiian crows (*Corvus hawaiiensis*).

PHOTO: SAN DIEGO ZOO WILDLIFE ALLIANCE

seek shelter. Establishing such correlations enables the formulation of hypotheses about signal function that can then be tested experimentally (e.g., with controlled playbacks).

Following this approach, decades of careful research have produced major advances in understanding animal communication (*3*). But there are considerable challenges, such as avoiding anthropocentric biases in data collection and interpretation, processing ever-increasing volumes of data, charting the full complexity of animals' signaling behavior, and achieving comprehensive functional decoding. ML offers some potential solutions.

Animal signals can be investigated using a rich toolkit of increasingly powerful ML methods, which vary in their modeling objectives, data requirements, and reliance on expert annotation. This includes, among other approaches, supervised learning (e.g., for determining which features accurately predict human-labeled signal types) and unsupervised and self-supervised learning (e.g., for discovering signal repertoires of an individual, group, or population).

Self-supervised deep-learning methods (*4*) are of interest because they require neither annotated datasets nor predefining features that are potentially relevant for communication. They are also the basis of "foundation models," which are capable of remarkable generalizations across tasks (*5*). For example, large language models that have been trained to predict the next word from a given sequence of words can subsequently be used to carry out much more complex tasks, such as inferring the syntactic classes of, and relations between, linguistic units, or generating realistic text (*5*).

Methods that can integrate different data modalities seem particularly promising for facilitating functional decoding because they can provide a fuller account of communication events. ML models have been developed that efficiently learn to link images to words, words to speech, and content across other modality combinations (*5*), and this approach could be applied productively to animal study systems, for example, by correlating vocalizations with specific behaviors. ML would effectively assist with the challenging task of detecting cross-modal associations (and structure) that can, in turn, inform the design of validation experiments to establish causality (see the figure).

Because many ML methods were originally developed for natural language processing, exciting avenues have started opening up for exploring the much-debated potential similarities between human language and animal communication systems (*6*). Observations and experimental work suggest that at least some animals, such as southern pied babblers (*Turdoides bicolor*), exhibit some of the order sensitivity and compositionality that are characteristic of human language (*7*). ML approaches could leverage large datasets to search for subtlety and complexity that elude traditional methods, potentially expanding the known set of communication features shared across divergent taxa.

There is a growing number of studies that are exploiting the potential of ML for investigating animal communication, including large collaborative initiatives, such as the Earth Species Project (ESP); Communication and Coordination Across Scales (CCAS); Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR); Interspecies Internet; and Project CETI (Cetacean Translation Initiative), which recently provided a detailed roadmap for ML-assisted work on sperm whale (*Physeter macrocephalus*) communication (*2*). Although efforts to tackle this grand research challenge are clearly intensifying, the field faces at least two main data-related obstacles: Most methods require vast amounts of data (*4*), and recordings of a single modality (e.g., vocalizations) are insufficient for functional decoding; additional context is required, including information on the animals' behavior and environment.

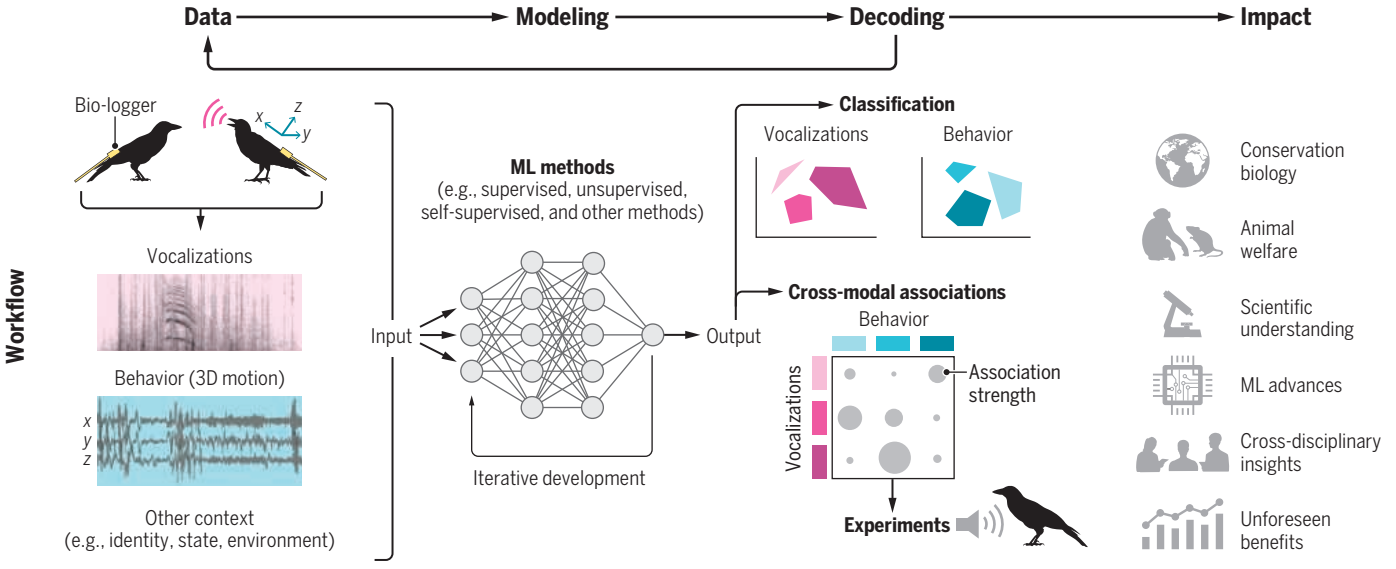
Large volumes of audio and video data are held in community-sourced archives (such as the Macaulay Library or xeno-canto), are being accumulated by passive recording arrays, or can be scraped from the internet. Mining these data sources will provide fascinating glimpses of the richness of animal communication, but on its own, such work is unlikely to achieve breakthroughs in decoding signal function. This is chiefly because robust information on the identities and states of the senders and receivers, and the specific communication context, is usually lacking.

High-quality datasets are available for some taxa, enabling swift progress with core model-development objectives. But it is clear that community mobilization and appropriate resourcing are required to ensure that species experts are fully involved in the annotation and interpretation of existing recordings and can lead targeted efforts to collect new data at scale, in both the laboratory and field. For wild animals, a range of methods can be used to collect suitable datasets, including observation of focal subjects, autonomous cameras and audio recorders, drones and robots, and animal wearables (bio-loggers). Some bio-logging devices can collect audio and body-motion data simultaneously for the same individual, providing valuable input for multimodal ML models (see the figure).

The journey could be the biggest reward. Training the lens of ML on a broad range of taxa will likely uncover surprising degrees

Using multimodal data and experiments to understand animal signals

Machine-learning (ML) methods can be used to integrate information on sender, receiver, and communication context, revealing patterns that may inform hypotheses about signal function and, in turn, the design of controlled experiments. ML-assisted research on animal communication will likely generate important benefits, such as improving animal conservation and welfare, but is not without its challenges; addressing ethical concerns is a top priority.



Challenges			
Collection of suitable data	Amount and quality of input data	Model interpretability	Effective stakeholder engagement
Broader context often unknown	Scarce data annotation by experts	Ethical concerns about experiments	Misuse of open data and code
Social complexity	Lack of suitable benchmarks	Access to suitable study systems	Translation of findings
Disturbance to wild animals	Computational resources required	Inconclusive results	Adequate community support

of previously hidden complexity in animals' communicative behavior. Many of the species that appear to use only a handful of basic call types may turn out to possess rich vocal repertoires, and those that are renowned for their sophisticated communication may be shown to be more impressive still. There are early indications of the discovery potential of ML, as highlighted by a recent study that explored individual and group differences in the vocal behavior of zebra finches (*Taeniopygia guttata*) (*8*).

The ability of ML to produce systematic inventories of vocal (or other signaling) output across a diverse range of taxa will enable unprecedented comparative analyses, helping researchers to pinpoint the evolutionary drivers, genomic signatures, life-history correlates, and cognitive and sensory foundations of different communication systems. At the same time, longitudinal recordings for individual subjects could reveal how communication skills arise and mature (*9*).

But perhaps most importantly, advances in this field could boost animal conservation and welfare. For example, in critically endangered species, such as the Hawaiian crow (*Corvus hawaiiensis*), comparisons with historical baseline data could generate a detailed record of how population bottlenecks have altered vocal repertoires, potentially leading to impoverished communicative capabilities (*10*); lost calls of high fitness relevance, such as those involved in foraging, courtship, or antipredator behavior, could then conceivably be reintroduced. Furthermore, there is increasing recognition that socially transmitted information may affect population viability (*11*), as illustrated by foraging specializations in killer whales (*Orcinus orca*) (*12*). Where vocal dialects can be established as "cultural markers," ML approaches would enable automated mapping of social population structure and identification of animal groups at risk of losing critical knowledge.

ML could also be used to identify animal signals that are associated with stress, discomfort, pain, and evasion, or with positive states, such as arousal and playfulness. This could provide momentum for improving the living conditions of livestock and other captive animals and may even enable the assaying of wild populations to measure the impact of anthropogenic stressors. Ecological "soundscape" analyses are, at present, largely focused on species detection, but it should be possible to listen in on animals' welfare at the landscape level (*13*). This idea could be developed further by looking beyond communication, for example, by developing ML tools that can examine satellite-recorded animal movement tracks for signatures of disease, distress, or human avoidance.

Despite manifold potential benefits, ML-assisted research on animal communication raises major ethical questions, such as under what circumstances it is acceptable to conduct playback experiments with wild animals. Advanced chatbots may enable researchers to initiate communication with animals before signal function is fully understood, potentially causing unintended harm. For example, broadcasting vocalizations to wild humpback whales (*Megaptera novaeangliae*) could inadvertently trigger changes in singing behavior on an ocean-basin scale. These issues must be tackled head-on and not as an afterthought. Cross-stakeholder consultation is urgently needed to develop best-practice guidelines and appropriate legislative frameworks (*14*).

Other challenges and opportunities lie ahead. For example, it is important to coordinate research efforts across existing initiatives and to enhance engagement of experts on animal communication, tracking, conservation, and welfare. Despite rapid technological advances, progress in this field will continue to depend on careful consideration of each study species' biology, a detailed knowledge of communication contexts, and controlled behavioral experiments (*3*). This expertise is essential for informing and validating ML analyses and intensifying data interpretation and collection efforts. Professional societies

and networks could help coordinate inclusive community-driven collaborations.

Workflows should be developed using study systems where data collection and experimental validation are relatively straightforward. In captive settings, researchers can ensure excellent experimental control as well as the highest ethical and welfare standards; good models include rodents, bats, and birds. Such work can be complemented with analyses of extensive field datasets that are already available for some species. Once methods have been established, they can be cautiously applied to the considerably more challenging problem of studying difficult-to-observe wild animals.

Present developments in ML are exceptionally fast paced. Beyond the use of deep-learning methods, there is scope for trialing other ML frameworks, such as reinforcement learning and meta-learning (i.e., learning from the output of other ML models). As models are developed, formal "benchmarking" will be key to improving the reliability and efficiency of analysis pipelines (*15*), although safeguards must be put in place to prevent the misuse of open resources, such as attempts to disturb, kill, or weaponize animals.

ML holds the potential to generate transformative advances in our understanding of animal communication systems, uncovering unimagined degrees of richness and sophistication. But it is essential that future advances are used to benefit the animals being studied. ■

REFERENCES AND NOTES

1. D. Tuia *et al.*, *Nat. Commun.* **13**, 792 (2022).
2. J. Andreas *et al.*, *iScience* **25**, 104393 (2022).
3. J. W. Bradbury, S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, 2011).
4. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
5. R. Bommasani *et al.*, arXiv:2108.07258 (2022).
6. A. R. Fishbein *et al.*, *Philos. Trans. R. Soc. London Ser. B* **375**, 20190042 (2020).
7. S. Engesser, S. W. Townsend, *WIREs Cogn. Sci.* **10**, e1493 (2019).
8. J. Goffinet *et al.*, *eLife* **10**, e67855 (2021).
9. B. C. Roy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12663 (2015).
10. A. M. Tanimoto *et al.*, *Anim. Behav.* **123**, 427 (2017).
11. P. Brakes *et al.*, *Science* **363**, 1032 (2019).
12. A. D. Foote *et al.*, *Nat. Commun.* **7**, 11693 (2016).
13. B. Williams *et al.*, *Ecol. Indic.* **140**, 108986 (2022).
14. O. R. Wearn, R. Freeman, D. M. P. Jacoby, *Nat. Mach. Intell.* **1**, 72 (2019).
15. M. Hagiwara *et al.*, in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2023), pp. 1–5.

ACKNOWLEDGMENTS

The authors thank V. Baglione, D. Canestrari, M. Cusimano, F. Effenberger, D. Gruber, M. Hagiwara, B. Hoffman, M. Johnson, S. Keen, I. Kuksa, J. Lawton, J. Y. Liu, M. Miron, R. Patchett, L. Rendell, R. Seyfarth, P. Shaw, and additional colleagues from ESP and Project CETI for discussion, comments, and support. Conflicts of interest and funding sources are listed in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7314

10.1126/science.adg7314

POLICY FORUM

Do we want less automation?

AI may provide a path to decrease inequality

By **Ajay Agrawal, Joshua S. Gans, Avi Goldfarb**

Impressive achievements made through artificial intelligence (AI) innovations in automating the tasks required in many jobs have reinforced concerns about labor market disruption and increased income inequality. This has motivated calls for change in the direction of AI innovation from being guided by task automation to instead focusing on labor augmentation (*1*). But task automation and labor augmentation are not polar opposites. Instead, automation of some tasks can lead to augmentation of labor elsewhere. Furthermore, AI automation may provide a path to reversing the trend of increasing income inequality by enabling disproportionate productivity improvements for lower-wage workers, allowing them to perform at levels that would previously require years of education and experience.

People have worried about automation—using machines to do the work that humans do—for centuries. Over that time, automation has proceeded apace, with entire sectors such as agriculture and manufacturing going from majority to minority shares of employment in many countries. Despite these sweeping changes, continual productivity improvements have not brought about technological unemployment. There has always been more for people to do, even if the fruits of economic growth have in recent times not lifted the incomes of all (*2*). Nonetheless, with the accelerated pace of innovations in AI technologies that specifically target the automation of cognitive rather than physical labor, many economists have become concerned that this time will be different, leading to substantial disruption and increased inequality but with little benefit to productivity and standard of living (*3*, *4*).

The economic history of the internet and computing over the past 50 years suggests that worries about inequality are not unfounded. These were what economists call skill-biased technologies (*5*), placing increasing demands on the skill, education, and know-how of the workforce. As computers and the internet diffused in the decades after 1980, demand for skills grew faster than supply, and inequality increased because of the disproportionate increase in wages for high-skilled workers.

More recently, consider all of the new products based on generative AI that automate the tasks of reading, writing, editing, summarizing, composing music, creating images, synthesizing speech, translating languages, programming computers, and producing videos. In each case, they promise time-saving productivity boosts by substituting capital (hardware or software) for the labor time that would otherwise be devoted to such tasks. If a machine can do these tasks, what will become of the people who did them previously? What if, instead of seeking to automate existing tasks, the mindset of the innovators was to provide tools that make existing workers more productive in their current jobs?

AUTOMATION AND AUGMENTATION

Automation and augmentation need not be opposites; the economic definitions of each are not presented as such. Economists studying automation consider a job to be a collection of tasks [as developed in (*4*, *6*) and applied in (*7*-*9*)]. Automation occurs when a machine (capital) is substituted for labor performance of one or more tasks in a workflow. This will typically increase productivity, which determines long-term economic growth and the average standard of living. With technological innovation having long been recognized as a key source of productivity growth, there is close to a consensus among leading economists that AI-driven technological change is likely to increase productivity, economic growth, and the average standard of living (*10*). However, if workers are replaced by machines, then this improved productivity can come at a substantial cost to displaced workers.

If technology creates new tasks, then displaced workers can find new things to do (*4*). Augmentation can therefore be seen as the development of new tasks (*3*), particularly if those tasks complement existing human labor (*1*). There is a challenge, however, in identifying such tasks and determining which inventions are likely to lead to their development. The difficulty is that one person's automation is another's augmentation. Automating one task may create even more new tasks.

Rotman School of Management, University of Toronto, Toronto, ON, Canada. Email: avi.goldfarb@rotman.utoronto.ca

and networks could help coordinate inclusive community-driven collaborations.

Workflows should be developed using study systems where data collection and experimental validation are relatively straightforward. In captive settings, researchers can ensure excellent experimental control as well as the highest ethical and welfare standards; good models include rodents, bats, and birds. Such work can be complemented with analyses of extensive field datasets that are already available for some species. Once methods have been established, they can be cautiously applied to the considerably more challenging problem of studying difficult-to-observe wild animals.

Present developments in ML are exceptionally fast paced. Beyond the use of deep-learning methods, there is scope for trialing other ML frameworks, such as reinforcement learning and meta-learning (i.e., learning from the output of other ML models). As models are developed, formal "benchmarking" will be key to improving the reliability and efficiency of analysis pipelines (*15*), although safeguards must be put in place to prevent the misuse of open resources, such as attempts to disturb, kill, or weaponize animals.

ML holds the potential to generate transformative advances in our understanding of animal communication systems, uncovering unimagined degrees of richness and sophistication. But it is essential that future advances are used to benefit the animals being studied. ■

REFERENCES AND NOTES

1. D. Tuia *et al.*, *Nat. Commun.* **13**, 792 (2022).
2. J. Andreas *et al.*, *iScience* **25**, 104393 (2022).
3. J. W. Bradbury, S. L. Vehrencamp, *Principles of Animal Communication* (Sinauer, 2011).
4. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
5. R. Bommasani *et al.*, arXiv:2108.07258 (2022).
6. A. R. Fishbein *et al.*, *Philos. Trans. R. Soc. London Ser. B* **375**, 20190042 (2020).
7. S. Engesser, S. W. Townsend, *WIREs Cogn. Sci.* **10**, e1493 (2019).
8. J. Goffinet *et al.*, *eLife* **10**, e67855 (2021).
9. B. C. Roy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12663 (2015).
10. A. M. Tanimoto *et al.*, *Anim. Behav.* **123**, 427 (2017).
11. P. Brakes *et al.*, *Science* **363**, 1032 (2019).
12. A. D. Foote *et al.*, *Nat. Commun.* **7**, 11693 (2016).
13. B. Williams *et al.*, *Ecol. Indic.* **140**, 108986 (2022).
14. O. R. Wearn, R. Freeman, D. M. P. Jacoby, *Nat. Mach. Intell.* **1**, 72 (2019).
15. M. Hagiwara *et al.*, in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2023), pp. 1–5.

ACKNOWLEDGMENTS

The authors thank V. Baglione, D. Canestrari, M. Cusimano, F. Effenberger, D. Gruber, M. Hagiwara, B. Hoffman, M. Johnson, S. Keen, I. Kuksa, J. Lawton, J. Y. Liu, M. Miron, R. Patchett, L. Rendell, R. Seyfarth, P. Shaw, and additional colleagues from ESP and Project CETI for discussion, comments, and support. Conflicts of interest and funding sources are listed in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7314

10.1126/science.adg7314

POLICY FORUM

Do we want less automation?

AI may provide a path to decrease inequality

By **Ajay Agrawal, Joshua S. Gans, Avi Goldfarb**

Impressive achievements made through artificial intelligence (AI) innovations in automating the tasks required in many jobs have reinforced concerns about labor market disruption and increased income inequality. This has motivated calls for change in the direction of AI innovation from being guided by task automation to instead focusing on labor augmentation (*1*). But task automation and labor augmentation are not polar opposites. Instead, automation of some tasks can lead to augmentation of labor elsewhere. Furthermore, AI automation may provide a path to reversing the trend of increasing income inequality by enabling disproportionate productivity improvements for lower-wage workers, allowing them to perform at levels that would previously require years of education and experience.

People have worried about automation—using machines to do the work that humans do—for centuries. Over that time, automation has proceeded apace, with entire sectors such as agriculture and manufacturing going from majority to minority shares of employment in many countries. Despite these sweeping changes, continual productivity improvements have not brought about technological unemployment. There has always been more for people to do, even if the fruits of economic growth have in recent times not lifted the incomes of all (*2*). Nonetheless, with the accelerated pace of innovations in AI technologies that specifically target the automation of cognitive rather than physical labor, many economists have become concerned that this time will be different, leading to substantial disruption and increased inequality but with little benefit to productivity and standard of living (*3*, *4*).

The economic history of the internet and computing over the past 50 years suggests that worries about inequality are not unfounded. These were what economists call skill-biased technologies (*5*), placing increasing demands on the skill, education, and know-how of the workforce. As computers and the internet diffused in the decades after 1980, demand for skills grew faster than supply, and inequality increased because of the disproportionate increase in wages for high-skilled workers.

More recently, consider all of the new products based on generative AI that automate the tasks of reading, writing, editing, summarizing, composing music, creating images, synthesizing speech, translating languages, programming computers, and producing videos. In each case, they promise time-saving productivity boosts by substituting capital (hardware or software) for the labor time that would otherwise be devoted to such tasks. If a machine can do these tasks, what will become of the people who did them previously? What if, instead of seeking to automate existing tasks, the mindset of the innovators was to provide tools that make existing workers more productive in their current jobs?

AUTOMATION AND AUGMENTATION

Automation and augmentation need not be opposites; the economic definitions of each are not presented as such. Economists studying automation consider a job to be a collection of tasks [as developed in (*4*, *6*) and applied in (*7*-*9*)]. Automation occurs when a machine (capital) is substituted for labor performance of one or more tasks in a workflow. This will typically increase productivity, which determines long-term economic growth and the average standard of living. With technological innovation having long been recognized as a key source of productivity growth, there is close to a consensus among leading economists that AI-driven technological change is likely to increase productivity, economic growth, and the average standard of living (*10*). However, if workers are replaced by machines, then this improved productivity can come at a substantial cost to displaced workers.

If technology creates new tasks, then displaced workers can find new things to do (*4*). Augmentation can therefore be seen as the development of new tasks (*3*), particularly if those tasks complement existing human labor (*1*). There is a challenge, however, in identifying such tasks and determining which inventions are likely to lead to their development. The difficulty is that one person's automation is another's augmentation. Automating one task may create even more new tasks.

Rotman School of Management, University of Toronto, Toronto, ON, Canada. Email: avi.goldfarb@rotman.utoronto.ca

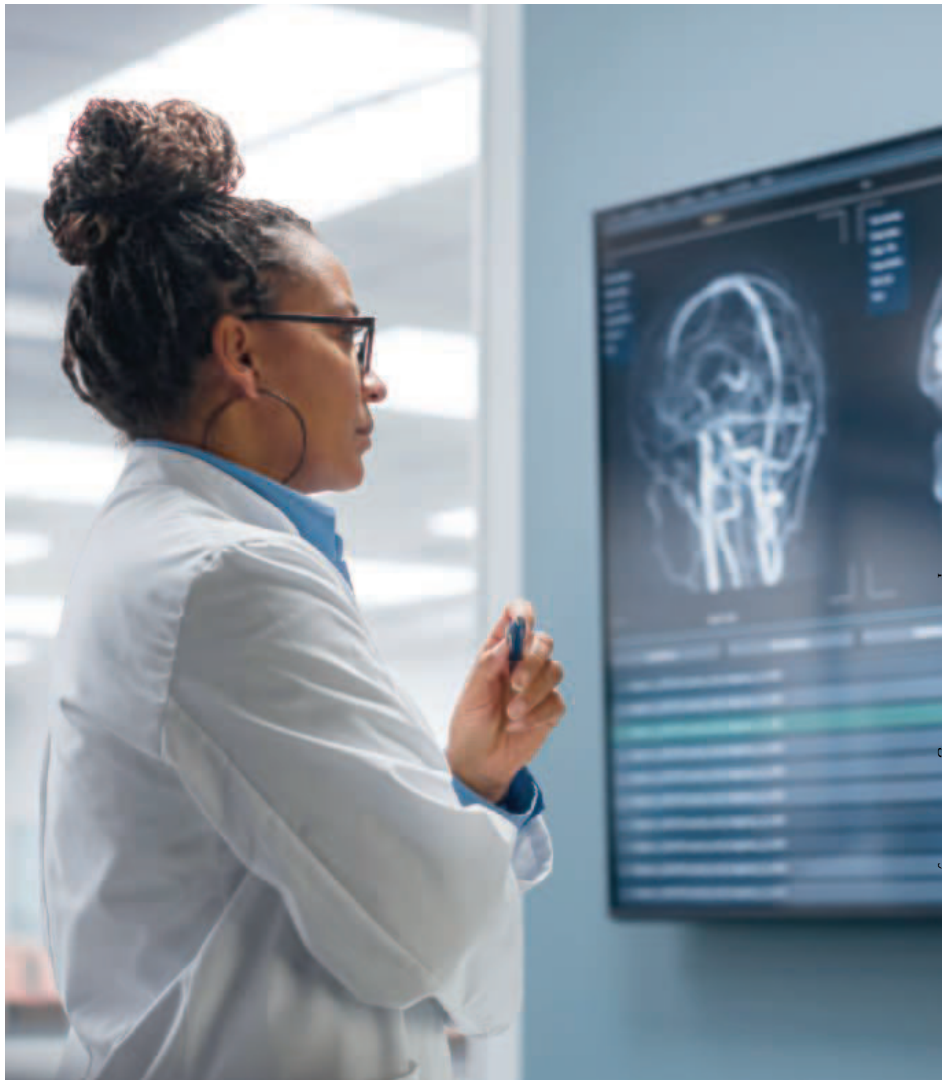
The use of a machine in a task is labor augmenting if the increase in productivity in that task raises labor productivity in other tasks. This will happen if tasks are complements. It may also arise if the time saved in performing the now automated task allows allocation of time to other tasks that involve higher productivity boosts. The impact on labor demand, in the end, depends on how much more productive the automated process is. Automation may be the correct decision for managers, yet in aggregate, the productivity benefit may be small (*4*). Several economists have argued that the current tax system and corporate environment encourage such incrementally useful automation ("so-so technologies") that substitutes for, rather than complements, labor (*3*, *4*).

This distinction between augmentation and automation is illustrated by a recent study of the impact of AI on Tokyo taxi drivers (*11*). The task in question was to predict what locations would have greater consumer demand for rides at any given time. A product, AI Navi, was designed with the goal of automating the task of predicting passenger demand. This improved the productivity of drivers, who spent up to 7% less time driving without passengers. Thus, one task—predicting passenger demand—was automated, which augmented the productivity of another task—driving—and with it, the taxi drivers' entire job.

AI Navi was a direct substitute for worker effort and skill. For experienced drivers, this substitution did not increase the productivity of that task, and hence, job performance was essentially unchanged. For less-skilled workers, however, their overall job performance improved by 14%. Similar results have been found in other contexts. For example, the use of a generative AI assistant in a call center substantially improved the productivity of novice and low-skilled workers but had little impact on the most productive workers (*12*).

This type of impact stands in contrast to the "skill-biased technical change"—when those with higher wages had their work augmented, and others had their work automated—that economists believe explains much of the increases in inequality over the past few decades. For example, increased adoption of computing was shown to drive expansion of high-wage and low-wage work at the expense of middle-wage jobs in the 1980s (*13*). This change was likely a result of computing complementing abstract high-education tasks and substituting routine middle-education tasks.

Often, the computer did not automate the core tasks of particular workers but augmented those with specific skills. Consider word processing. Before word processing on personal computers, typing was a separate process that involved a specialized skill that was central to millions of jobs. A typist did several related tasks involving typing, error correction, and page setting. Word processing on personal computers did not automate typing, but it did automate error correction (the entire document no longer needed to be retyped). This made it efficient for knowledge workers to type themselves because automating the time-consuming task of correcting errors meant that typing errors were no longer consequential. This augmented what a knowledge worker could do. In the process, it reduced demand for secretarial workers who received midlevel wages and increased demand for knowledge workers who already were earning high wages. To economists, automated error correction may have seemed like augmentation because it is not listed as a specific task in the databases they use to understand workflows [see supplementary materials (SM)]. Instead, for typists, secretaries, or administrative assistants, automated error correction is an implied part of other tasks such as checking for grammar and formatting errors, which did not get automated.



More generally, computers replaced workers engaged in routine technical duties, such as bookkeeping, administrative tasks, and manufacturing processes. Simultaneously, they enhanced the abilities of educated professionals who excel in areas such as problem-solving, creativity, and persuasion (*6*).

HOW AI MIGHT REDUCE INEQUALITY

When we look further into AI success stories of automation, it is often the case that tasks are automated in a way that is not skill biased, and indeed is the opposite: allowing more people to perform entire jobs at a level that previously required expensive skills. Hence, they present an opportunity to reduce the premium paid to workers with those skills and distribute employment opportunities more widely.

AI-driven diagnostic tools, with appropriate policy support, hold potential to expand opportunities and boost wages for nonphysician health care workers.

Diagnosis

A burgeoning body of research has demonstrated the potential of AI in delivering medical diagnoses (*14*). At the core of this research lies the understanding that diagnosis is fundamentally a predictive process, taking known symptoms and determining their underlying causes. Traditionally, diagnostic expertise has been the cornerstone of a physician's skillset, with medical schools and selection processes emphasizing accurate diagnosis.

Whereas other health care professionals may excel in patient support, stress management, or day-to-day care, the ability to diagnose is a key skill that sets doctors apart. Consequently, we can consider diagnosis as one task, with other medical responsibilities representing other tasks. The expertise required for diagnosis is relatively rare compared with the skills associated with other medical tasks.

AI-driven diagnostic tools, however, have the potential to automate this rare skill. In 2021, there were approximately 760,000 physician and surgeon positions in the United States, with median annual incomes exceeding $200,000 (see SM). By automating the core diagnostic skill that underpins these roles, AI has the potential to substantially disrupt the value that doctors bring to health care, potentially causing employment stagnation and wage suppression. This concern aligns with cautionary predictions regarding AI's replication of human intelligence (*1*, *3*).

By contrast, the wider health care profession—which includes 3 million registered nurses and millions of other professionals such as pharmacists, nurse practitioners, physician assistants, and paramedics—could potentially benefit from AI-driven diagnostics. Diagnostic capabilities often act as a barrier for these professionals to fully use their skills (*14*). By granting access to AI-assisted diagnostics, these nonphysician health care workers could experience an expansion in career opportunities and a potential boost in wages.

In this case, whether AI reduces inequality is not driven by a decision of whether to automate or augment. Instead, the impact depends on whose core skills get automated. If AI automates the higher-wage workers such as doctors, then it might augment other medical professionals and decrease inequality. By contrast, if the regulatory environment or industry incentive structure direct innovation toward automating the lower-wage workers—for example, because the decisions made by physicians cannot be delegated to others by law—then any AI automation will focus on augmenting those who are already earning high incomes and automating others (*14*).

Writing and editing

Generative AI tools such as ChatGPT suggest the potential to automate reading, writing, analyzing, and editing. Data from labor market analytics firm Lightcast show that 474,873 US job postings in January 2023 required writing or editing skills, out of a total of 3.4 million job postings (see SM). Thus, more than 10% of posted jobs are likely to be affected by automated writing tools. Jobs that require writing and editing skills have a higher posted salary on average than that for jobs that do not require such skills. The four job categories for which more than 75% of postings demand writing and editing skills are proofreaders and copy markers, technical writers, editors, and writers and authors. Without the writing and editing tasks that can be done with large language models (LLMs) such as ChatGPT, the jobs of proofreaders, copy markers, and technical writers might be completely replaced.

But the situation for editors, writers, and authors is different. There are a number of decisions involved that might not be automated with the recent advances in LLMs. In the presence of LLMs, the task that we classify as writing can be broken down into three tasks: prompting, writing, and signing off. Only the writing task is automated. To determine whether a job that previously assigned importance to writing skills would be affected positively or negatively, we would need to know how precisely those skills are now applied.

To see this, consider journalism, a job category in which 65% of job postings list writing or editing skills [see models in (*10*) and (*14*)]. Print journalism places high weight on writing skills. What is less obvious is whether journalists engage in prompting or signing off as part of their job. There are two possibilities. In job one, a journalist may be assigned an article to write by an editor. They then write the article, and the editor decides whether it can be published. In job two, a journalist comes to an editor with a proposal. They then write the article, and the editor decides whether it can be published. The only difference between these two jobs is who comes up with the idea for the article.

That distinction matters when we introduce LLMs that can do the narrowly defined writing task. If the editor is the prompter, the editor may prompt an LLM rather than a human journalist to do the writing. Thus, journalists who are not prompters may be less valued when LLMs are deployed. By contrast, if the journalist is the prompter, the LLM becomes a tool for the journalist to more efficiently produce the article. The LLM increases the value of what the journalist is doing in terms of identifying interesting topics to write about. At present, both jobs are classified as journalists, so we cannot determine how the entire class of workers will be affected. Furthermore, LLMs create opportunities for any people who have a good sense of which stories are valuable and important but who lack writing skills.

In each case, these are examples of de-skilling automation. Technology automates a task that requires highly specialized and highly paid workers, enabling less-specialized or lower-paid workers to include that task in their workflow. The demand for the skills of the highly-paid workers declines, whereas the demand for less-highly-paid workers is likely to increase. This should reduce inequality while increasing overall output, albeit reduc-

ing the relative (and perhaps absolute) income of the previously highly-paid workers whose skills have been automated.

It is too early to determine with much certainty how this will play out for AI, whether the impact on any particular job will be positive or negative. Research is beginning to emphasize which jobs are most likely to be affected rather than lost (*8*). For example, that classification tasks such as image recognition can be done with AI will affect workers whose jobs involve classification tasks, such as radiologists (*7*). Recent work examining differences between generative AI (specifically, LLMs) and nongenerative AI [as described in (*7*)] shows that millions of jobs have the potential to be affected by LLMs. Notably, these studies emphasize that "affect" does not mean "replace." For many jobs, automating some aspects of the workflow might increase productivity, the wages of workers who have that job, and the number of workers hired to do that job.

Even when some jobs get automated, that might complement the tasks done by other workers. Many empirical exercises [for example, (*7*–*9*)] emphasize the direct impact on jobs, but they do not explore the jobs that might be enhanced through complementary production processes. For example, in January 2023, there were 186,417 job postings in the United States that specified language skills (such as Spanish Language or American Sign Language), or about 5% of the total job postings (see SM). Automating language translation would directly affect many of these jobs. At the same time, many other jobs that do not require language skills would also be affected. For example, a recent study showed that small businesses that used a rudimentary automated language translation tool on eBay experienced a 17.5% increase in exports to markets where that language is used (*15*). Automation of some jobs could create opportunities for those whose work would appear to be unaffected, as measured with the tasks and skills involved in current workflows.

CONCLUSION

Many economists who have studied the impact of automation on labor markets have argued recently that the direction of AI research needs to be changed away from automating tasks to focusing on overall job augmentation. The implicit argument is that a focus on augmentation will lead to more complementarity with lower-wage labor and more new tasks. However, many recent advances in AI that have been developed with the explicit goal of task automation have appeared to increase worker productivity; that is, task automation has been labor augmenting. Furthermore, AI technology may disproportionately augment lower-skilled labor, reducing income inequality. This, at the very least, calls into question whether a change in the innovator's mindset is needed: Task automation may be a path to substantially improved labor productivity.

This potential to reverse the recent trend toward skill-biased technical change does not mean that AI is without risk. Other concerns remain, including those related to privacy, liberty, democracy, and monopoly power (*3*). Our emphasis is on understanding that one person's automation is another's augmentation, and that it is difficult for engineers or policy-makers to pick which particular innovation will increase or reduce inequality overall. We believe that both regulators and engineers should be careful in shutting down a particular technology trajectory because it appears to automate human work. In the process of automating some work, other work can be augmented.

Often, our analysis suggests that such augmentation from AI will increase the job productivity of less-skilled workers who can now perform at levels achieved by their skilled counterparts. This suggests that skill premia that have contributed to widening inequality may be eroded. Thus, it is quite plausible that the use of AI to automate tasks will both increase productivity and decrease income inequality. If so, then we may want more automation, not less. ■

REFERENCES AND NOTES

1. E. Brynjolfsson, *Daedalus* **151**, 272 (2022).
2. J. S. Gans, A. Leigh, *Innovation + Equality: Creating a Future that is more Star Trek than Terminator* (MIT Press, 2019).
3. D. Acemoglu, S. Johnson, *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity* (PublicAffairs, 2023).
4. D. Acemoglu, P. Restrepo, *Am. Econ. Rev.* **108**, 1488 (2018).
5. C. Goldin, L. Katz, *The Race Between Education and Technology* (Harvard Univ. Press, 2008).
6. D. H. Autor, *Science* **344**, 843 (2014).
7. E. Brynjolfsson, T. Mitchell, *Science* **358**, 1530 (2017).
8. E. Felten, M. Raj, R. Seamans, *Strateg. Manage. J.* **42**, 2195 (2021).
9. T. Eloundou, S. Manning, P. Mishkin, D. Rock, arXiv:2303.10130 [econ.GN] (2023).
10. A. Agrawal, J. S. Gans, A. Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (Univ. Chicago Press, 2019).
11. K. Kanazawa, D. Kawaguchi, H. Shigeoka, Y. Watanabe, "AI, skill, and productivity: The case of taxi drivers," no. w30612, National Bureau of Economic Research (2022).
12. E. Brynjolfsson, D. Li, L. Raymond, "Generative AI at work," National Bureau of Economic Research (NBER) working paper 31161 (NBER, 2023).
13. D. H. Autor, L. Katz, M. Kearney, *Rev. Econ. Stat.* **90**, 300 (2008).
14. A. Agrawal, J. S. Gans, A. Goldfarb, *Power and Prediction* (Harvard Business Review Press, 2022).
15. E. Brynjolfsson, X. Hui, M. Liu, *Manage. Sci.* **65**, 5449 (2019).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh9429

10.1126/science.adh9429

POLICY FORUM

Generative AI meets copyright

Ongoing lawsuits could affect everyone who uses generative AI

By **Pamela Samuelson**

Generative artificial intelligence (AI) is a disruptive technology that is widely adopted by members of the general public as well as scientists and technologists who are enthusiastic about the potential to accelerate research in a wide variety of fields. But some professional artists, writers, and programmers fiercely object to the use of their creations as training data for generative AI systems and to outputs that may compete with or displace their works (*1*, *2*). Lack of attribution and compensation for use of their original creations are other sources of aggravation to critics of generative AI. Copyright lawsuits that are now underway in the United States have substantial implications for the future of generative AI systems. If the plaintiffs prevail, the only generative AI systems that may be lawful in the United States would be those trained on public domain works or under licenses, which will affect everyone who deploys generative AI, integrates it into their products, and uses it for scientific research.

Conflicts between creators of copyrighted works and developers of technologies that enable the use of those creations in unexpected and innovative ways is nothing new. In the early 20th century, the disruptive technology of the day was player pianos. Music copyright owners sued the makers of piano rolls, claiming that rolls of their musical compositions were infringements. Subsequent copyright-disruptive technologies have included cable television, photocopiers, videotape recording machines, and MP3 players, each of which (except photocopiers) attracted copyright industry challenges (all of which failed in the courts, although Congress sometimes later extended protections in the aftermath of failed lawsuits).

Berkeley Law School, University of California Berkeley, Berkeley, CA, USA. Email: psamuelson@berkeley.edu

ing the relative (and perhaps absolute) income of the previously highly-paid workers whose skills have been automated.

It is too early to determine with much certainty how this will play out for AI, whether the impact on any particular job will be positive or negative. Research is beginning to emphasize which jobs are most likely to be affected rather than lost (*8*). For example, that classification tasks such as image recognition can be done with AI will affect workers whose jobs involve classification tasks, such as radiologists (*7*). Recent work examining differences between generative AI (specifically, LLMs) and nongenerative AI [as described in (*7*)] shows that millions of jobs have the potential to be affected by LLMs. Notably, these studies emphasize that "affect" does not mean "replace." For many jobs, automating some aspects of the workflow might increase productivity, the wages of workers who have that job, and the number of workers hired to do that job.

Even when some jobs get automated, that might complement the tasks done by other workers. Many empirical exercises [for example, (*7*–*9*)] emphasize the direct impact on jobs, but they do not explore the jobs that might be enhanced through complementary production processes. For example, in January 2023, there were 186,417 job postings in the United States that specified language skills (such as Spanish Language or American Sign Language), or about 5% of the total job postings (see SM). Automating language translation would directly affect many of these jobs. At the same time, many other jobs that do not require language skills would also be affected. For example, a recent study showed that small businesses that used a rudimentary automated language translation tool on eBay experienced a 17.5% increase in exports to markets where that language is used (*15*). Automation of some jobs could create opportunities for those whose work would appear to be unaffected, as measured with the tasks and skills involved in current workflows.

CONCLUSION

Many economists who have studied the impact of automation on labor markets have argued recently that the direction of AI research needs to be changed away from automating tasks to focusing on overall job augmentation. The implicit argument is that a focus on augmentation will lead to more complementarity with lower-wage labor and more new tasks. However, many recent advances in AI that have been developed with the explicit goal of task automation have appeared to increase worker productivity; that is, task automation has been labor augmenting. Furthermore, AI technology may disproportionately augment lower-skilled labor, reducing income inequality. This, at the very least, calls into question whether a change in the innovator's mindset is needed: Task automation may be a path to substantially improved labor productivity.

This potential to reverse the recent trend toward skill-biased technical change does not mean that AI is without risk. Other concerns remain, including those related to privacy, liberty, democracy, and monopoly power (*3*). Our emphasis is on understanding that one person's automation is another's augmentation, and that it is difficult for engineers or policy-makers to pick which particular innovation will increase or reduce inequality overall. We believe that both regulators and engineers should be careful in shutting down a particular technology trajectory because it appears to automate human work. In the process of automating some work, other work can be augmented.

Often, our analysis suggests that such augmentation from AI will increase the job productivity of less-skilled workers who can now perform at levels achieved by their skilled counterparts. This suggests that skill premia that have contributed to widening inequality may be eroded. Thus, it is quite plausible that the use of AI to automate tasks will both increase productivity and decrease income inequality. If so, then we may want more automation, not less. ■

REFERENCES AND NOTES

1. E. Brynjolfsson, *Daedalus* **151**, 272 (2022).
2. J. S. Gans, A. Leigh, *Innovation + Equality: Creating a Future that is more Star Trek than Terminator* (MIT Press, 2019).
3. D. Acemoglu, S. Johnson, *Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity* (PublicAffairs, 2023).
4. D. Acemoglu, P. Restrepo, *Am. Econ. Rev.* **108**, 1488 (2018).
5. C. Goldin, L. Katz, *The Race Between Education and Technology* (Harvard Univ. Press, 2008).
6. D. H. Autor, *Science* **344**, 843 (2014).
7. E. Brynjolfsson, T. Mitchell, *Science* **358**, 1530 (2017).
8. E. Felten, M. Raj, R. Seamans, *Strateg. Manage. J.* **42**, 2195 (2021).
9. T. Eloundou, S. Manning, P. Mishkin, D. Rock, arXiv:2303.10130 [econ.GN] (2023).
10. A. Agrawal, J. S. Gans, A. Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (Univ. Chicago Press, 2019).
11. K. Kanazawa, D. Kawaguchi, H. Shigeoka, Y. Watanabe, "AI, skill, and productivity: The case of taxi drivers," no. w30612, National Bureau of Economic Research (2022).
12. E. Brynjolfsson, D. Li, L. Raymond, "Generative AI at work," National Bureau of Economic Research (NBER) working paper 31161 (NBER, 2023).
13. D. H. Autor, L. Katz, M. Kearney, *Rev. Econ. Stat.* **90**, 300 (2008).
14. A. Agrawal, J. S. Gans, A. Goldfarb, *Power and Prediction* (Harvard Business Review Press, 2022).
15. E. Brynjolfsson, X. Hui, M. Liu, *Manage. Sci.* **65**, 5449 (2019).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh9429

10.1126/science.adh9429

POLICY FORUM

Generative AI meets copyright

Ongoing lawsuits could affect everyone who uses generative AI

By **Pamela Samuelson**

Generative artificial intelligence (AI) is a disruptive technology that is widely adopted by members of the general public as well as scientists and technologists who are enthusiastic about the potential to accelerate research in a wide variety of fields. But some professional artists, writers, and programmers fiercely object to the use of their creations as training data for generative AI systems and to outputs that may compete with or displace their works (*1*, *2*). Lack of attribution and compensation for use of their original creations are other sources of aggravation to critics of generative AI. Copyright lawsuits that are now underway in the United States have substantial implications for the future of generative AI systems. If the plaintiffs prevail, the only generative AI systems that may be lawful in the United States would be those trained on public domain works or under licenses, which will affect everyone who deploys generative AI, integrates it into their products, and uses it for scientific research.

Conflicts between creators of copyrighted works and developers of technologies that enable the use of those creations in unexpected and innovative ways is nothing new. In the early 20th century, the disruptive technology of the day was player pianos. Music copyright owners sued the makers of piano rolls, claiming that rolls of their musical compositions were infringements. Subsequent copyright-disruptive technologies have included cable television, photocopiers, videotape recording machines, and MP3 players, each of which (except photocopiers) attracted copyright industry challenges (all of which failed in the courts, although Congress sometimes later extended protections in the aftermath of failed lawsuits).

Berkeley Law School, University of California Berkeley, Berkeley, CA, USA. Email: psamuelson@berkeley.edu

When new technologies pose new copyright questions that Congress did not anticipate, courts typically consider which outcome is most consistent with the constitutional purposes of copyright. The Constitution gives Congress the power "to promote the progress of Science and useful Arts," that is, to foster the creation and dissemination of knowledge for the public good. This requires balancing the legitimate interests of copyright owners to prevent misappropriations of their works that undermine incentives to create with the legitimate interests of developers of innovative technologies and follow-on creators who need some breathing space in which they, too, can innovate.

What makes generative AI more disruptive than previous technologies? One factor is certainly the exceptionally rapid pace at which generative AI technologies have been launched, adopted, and adapted. Evolution in the fields of law and policy, by contrast and of necessity, is much slower. It is, moreover, not easy to assess how to calibrate balances among competing copyright interests in the early stages of new-technology evolutions. Generative AI seems poised to have substantial impacts on the careers of professional writers and artists. During the 2023 Writers Guild of America strike, for instance, uses of generative AI are one focus of negotiations. Screenwriters are understandably worried that these technologies will displace them or diminish their compensation.

Stability AI is defending two copyright infringement lawsuits in the United States that are focused on Stable Diffusion, a widely used image generator. Getty Images is the plaintiff in one of these lawsuits. The other is a class-action lawsuit on behalf of visual artists on whose images Stable Diffusion was trained. Both complaints assert that Stability AI made unlawful copies of the plaintiffs' images when ingesting them as inputs for training Stable Diffusion's model and that output images produced by Stable Diffusion in response to user prompts are infringing derivative works.

A third generative AI lawsuit (*Doe* v. *Github, Inc.*) challenges OpenAI's development of Codex, a large language model (LLM) trained upon billions of lines of open-source software code. Also challenged is GitHub and OpenAI's collaborative development of Copilot, a coding assistant tool that draws upon the Codex LLM to suggest lines of code for specific functions in response to user prompts. (Microsoft, which owns GitHub and has invested heavily in OpenAI, is a fellow defendant.)

Rulings in favor of plaintiffs might trigger "innovation arbitrage," causing developers of generative AI systems to move their bases of operation to countries that regard the ingestion of copyrighted works as training data as fair use, like Israel's Ministry of Justice did in early 2023. Other countries that want to attract AI innovations may follow suit. If courts uphold the *Stability AI* plaintiffs' claims, OpenAI's GPT4 and Google's BARD may also be in jeopardy. Their developers would be very attractive targets of follow-on lawsuits.

INGESTING TRAINING DATA

Stability AI has yet to articulate its main defenses to the copyright charges. Insofar as the complaints allege that Stable Diffusion contains copies of in-copyright images used as training data, the claims are factually and technically inaccurate. Stable Diffusion contains an extremely large number of parameters that mathematically represent concepts embodied in the training data, but the images as such are not embodied in its model.

Training a model begins by tokenizing the contents of works ingested as training data into component elements. The model uses these tokens to discern statistical correlations—often at staggeringly large scales—among features of the content on which the model is being trained. In essence, the model is extracting and analyzing precise facts about, and correlations between, discrete elements of the works to ascertain which other discrete elements either do or do not follow or are proximate to these elements and the frequency with which the correlations do or do not exist in varying contexts.

"The complaints against Stability AI overlook the intentionally porous nature of copyrights."

The complaints against Stability AI overlook the intentionally porous nature of copyrights. What copyright law protects is only the original expression that authors contribute (such as sequences of words in a poem or the melody of music). Copyright's scope never extends to any ideas, facts, or methods embodied in works nor to elements common in works of that kind (under copyright's "scenes a faire" doctrine), elements capable of being expressed in very few ways (under the "merger" doctrine), or the underlying subjects depicted in protected works. Photographs of cats, for instance, do not give the photographer exclusive rights to characteristic features of cats, such as their noses or facial expressions. Nor does copyright's scope extend to inferences that readers might draw from reviewing an author's works, such as insights about patterns of connections among concepts or how works of that kind are constructed.

Moreover, Stability AI did not prepare the dataset on which the Stable Diffusion model was trained. This was done by a nonprofit German research organization known as LAION (Large-Scale Artificial Intelligence Open Network). LAION initially developed LAION-5B, a dataset consisting of 5.85 billion hyperlinks that pair images and text descriptions from the open internet. LAION makes this dataset available to the public for free for use as training data for those who want to use it to build generative models. LAION also developed a subset of LAION-5B, known as LAION-Aesthetics, that consists of hyperlinks to 600 million images selected by some human testers for their visual appeal and by a machine-learning analysis of human aesthetic ratings. The Stable Diffusion model was trained on the LAION-Aesthetics dataset.

LAION's creation of this dataset was very likely lawful because the European Union (EU) adopted an exemption allowing nonprofit research organizations to make copies of in-copyright works for text and data mining (TDM) purposes. The EU created this exception in recognition of the societal value of TDM as a means by which researchers can create new knowledge. This exemption cannot be overridden by contract. (A second EU exemption authorizes commercial actors to engage in TDM, although copyright owners can opt out of this exemption, as some have done.)

Stability AI makes Stable Diffusion available on an open-source basis. However, it also provides a subscription service so that those who lack resources or the inclination to host the open-source version can have access to Stable Diffusion to generate images in response to text prompts. Yet, insofar as ingesting in-copyright images to train a generative model requires making at least temporary or incidental copies of them, Stability AI is likely to argue that this is a fair use under US copyright law.

FAIR USE

Under US law, fair uses of in-copyright works do not infringe copyrights. Courts consider four factors when assessing fair use defenses: (i) the purpose of the challenged use, (ii) the nature of the copyrighted works, (iii) the amount and substantiality of the taking, and (iv) the effect of the challenged use on the market for or value of the

copyrighted work. The purpose and market effects factors are generally the most important determinants in fair-use cases, but all four factors must be weighed together in a holistic analysis.

Research, scholarship, and teaching are among the favored fair-use purposes, as are criticism, comment, and news reporting. Noncommercial uses are generally favored more than commercial uses. Since 1994, when the Supreme Court considered the fairness of 2 Live Crew's rap parody of a popular Roy Orbison song in *Campbell* v. *Acuff-Rose Music, Inc.*, courts have given considerable weight to whether the purpose of a challenged use was "transformative." The Court defined this term as uses that "add[] something new, with a further purpose or different character, altering the first with new expression, meaning, or message" [(*3*), p. 579]. Transformative uses are also less likely than nontransformative uses to harm the market for the first work. People are, for instance, unlikely to purchase 2 Live Crew's parody if they want to listen to Roy Orbison's rendition.

The *Stability AI* plaintiffs will likely argue that the ingestion of their works as training data was nontransformative and commercial. Both considerations would, if accepted, tip against fair use. However, several court decisions have ruled that analogous digital uses of in-copyright works qualified as transformative fair uses.

For example, in *Authors Guild* v. *Google, Inc.*, a court ruled that Google's digitization of millions of books from research library collections to index their contents and serve a few snippets of book contents in response to user search queries was a "highly transformative" fair use. Although Google's purpose was commercial, it was very different from the purposes for which the books were marketed. Google's use facilitated greater public access to knowledge as well as enabling TDM research and the creation of new research tools. In *Field* v. *Google, Inc.*, a court found that Google's cache copying of contents from Field's website was a transformative fair use.

The nature-of-the-work factor often has little importance in fair-use cases. The *Stability AI* plaintiffs may argue that because works of visual art lie at the very core of copyright, fair use should be thinner for these works than for the old library books at issue in the *Authors Guild* case. A countervailing consideration is that the visual artists on whose works Stable Diffusion trained made their works available on the open internet, as did Field in the *Google, Inc.*, case.

Transformative purposes tend to have spillover effects on other fair-use factors, especially the amount factor. As in the *Authors Guild* case, the *Stability AI* plaintiffs may emphasize that the defendant made exact copies of the entirety of many millions of works without permission or compensation. However, courts typically inquire whether such copying was necessary to achieve a transformative purpose. In the *Authors Guild* case, the court recognized that Google could not index book contents and serve up snippets in response to search queries unless it copied the books' contents. Stability AI will likely make a similar necessity argument about training-data usages of images.

The market effect of a challenged use is sometimes said to be the most important fair-use factor. The Getty complaint against Stablity AI emphasizes that it has established a licensing market for use of its premium photographs as training data for generative AI. That bolsters Getty's argument that Stability AI's appropriation of 12 million images from Getty websites has harmed a licensing market. The class-action claim against Stability AI is weaker because Stability AI could not have gotten a license from the class of visual artists whose works were ingested to construct the Stable Diffusion model.

The existence of a licensing market (or an intent to establish one) is not, however, a consideration that by itself can resolve a dispute in transformative fair use cases. In its 2021 *Google LLC* v. *Oracle America, Inc.*, decision, the Supreme Court rejected Oracle's argument that Google's use of parts of the Java application programming interface (API) had deprived Oracle of license revenues to which it claimed an entitlement. The Court stated that courts should consider the public benefits of a challenged use as well as potential lost revenues and how much creativity a challenged use has enabled and balance this against potential losses.

This consideration was very relevant in the *Oracle* case. Not only was Google's Android smartphone platform, in which the Java API was used, a highly innovative new software product, but it enabled millions of programmers to use their familiarity with the Java API to create many millions of programs. The Court thought this use was consistent with the constitutional objective of copyright to promote creative progress. The public greatly benefited from Android's existence and the availability of large numbers of apps that ran on that platform.

Stability AI will almost certainly channel

the public-benefit and creative-impacts statements in the *Oracle* decision and point to the exceptional creativity embodied in Stable Diffusion as well as the hundreds of millions of creative uses of this generative AI system, including those by graphic artists who use it to generate ideas or refine creations.

The *Stability AI* plaintiffs will likely counter this argument with the Supreme Court's 2023 ruling in *Andy Warhol Foundation for the Visual Arts, Inc.* v. *Goldsmith*, which somewhat narrowed the conception of transformative purposes. It no longer suffices for challenged works to have a new meaning or message. More important now are whether the challenged use has a different purpose than the first work and how commercial the use is. Stability AI will argue that ingesting copyrighted materials as training data had a very different purpose than the works as first published.

What might tip the scale against Stability AI's fair-use defense is whether images produced by Stable Diffusion infringe the derivative work right of the authors of the images on which its LLM was trained. A relevant precedent is *Sega Enterprises Ltd.* v. *Accolade, Inc.*, in which an appellate court decided that Accolade had made fair use of Sega software when making reverse-engineered copies for the legitimate purpose of extracting information about how to make its videogames compatible with the Sega platform. Had Accolade reverse-engineered for an illegitimate purpose, such as to appropriate expression from the Sega games, its fair-use defense would have faltered. The Accolade games competed with Sega's games, but the court thought that this was the kind of competition among noninfringing works that copyright is supposed to foster.

OUTPUTS AS INFRINGEMENTS

The class-action complaint against Stability AI asserts that all images produced by Stable Diffusion are infringing derivative works because all are derived from the images on which its model trained. It characterizes Stable Diffusion as a "collage tool" whose outputs compete against the artists' own works and thereby harm their markets. Users of Stable Diffusion, moreover, can submit prompts requesting the generation of an image of a particular subject "in the style of" a specific named artist.

However, courts have long held that to infringe copyright's derivative work right, it is not enough to show that a second work was "based upon" an earlier work or some of its elements. The second work must have appropriated a substantial quantum of the first work's original expression. So, unless a court decides to overturn decades of precedents interpreting the derivative work right and broaden it substantially, the class action's output infringement claim is likely to fail.

The class-action complaint acknowledges that "[i]n general, none of the Stable Diffusion output images provided in response to a particular Text Prompt is likely to be a close match for any specific image in the training data" [(*4*), p. 23]. Even "in the style of" claims seem weak because copyright law does not protect styles as such. Infringement can be found only if there is a close resemblance between expressive elements of a stylistically similar work and original expression in particular works by that artist.

The reason that Stable Diffusion outputs are highly unlikely to be substantially similar to particular images on which its model was trained is due to how Stable Diffusion assembles them. Constructing a model for an image-generating AI requires processing enormous quantities of input data to produce abstract representations of image elements (such as cats playing with a ball on a linoleum floor). Diffusion adds noise to image elements when encoding them. The pairing of text descriptions and images allows the model to cluster the abstract representations so that similar representations will be in proximity (representations of cats near other cat representations). When a user enters a prompt directing the software to generate a specific type of output, the generative AI system uses complex statistical calculations to assemble an output that the system predicts will match what the user requested.

It is, however, possible for generative AI outputs to infringe copyrights. If the same input image (say, of Mickey Mouse) is present in many works on which the model was trained and its developer did not follow industry best practices by eliminating duplicates and using output filters to prevent infringements, user prompts could result in infringing outputs (although this user, not the developer of the generative AI system, may be the infringer). Ironically, the larger and more diverse that the dataset on which a generative model was trained is, the less likely are infringing outputs.

The Getty complaint against Stability AI is more modest in its infringing output claims. Yet Getty, too, may find it difficult to prove that particular Stable Diffusion outputs are substantially similar to particular photographs to which it owns copyrights. In general, Stable Diffusion outputs will be distinguishably different from the images on which the model was trained.

The *Stability AI* plaintiffs will likely emphasize that the images produced by Stable Diffusion compete with their works in the marketplace. They can point to the Supreme Court's *Goldsmith* decision, which treated competing uses as weighing against fairness. Yet *Goldsmith* involved two works that were substantially similar in their expressions—Goldsmith's photo of Prince and Warhol's print derived from Goldsmith's photo—that competed in the same licensing market for magazines. Stability AI will be relying on differences in Stable Diffusion's outputs relative to plaintiffs' works in order to distinguish their case's context from that of *Goldsmith*.

CONCLUDING THOUGHTS

Based on existing precedents and an understanding about how Stable Diffusion was trained and how it generates images in response to prompts, Stability AI seemingly has a reasonable chance of prevailing on the copyright claims. (Both the Getty and the class-action complaints raise other claims that cannot be addressed in this brief article.) The lawsuits are, however, in very early stages, and it may be years before courts render decisions.

In mid-May 2023, Congress held its first hearing about generative AI and copyright issues, during which witnesses expressed divergent views. The US Copyright Office is well aware of the consternation that generative AI has fomented in copyright-dependent communities. The Office hosted "listening sessions" in spring 2023 to provide stakeholders with opportunities to explain their perspectives on the two principal questions posed in the *Stability AI* cases: Is the use of in-copyright works as training data for generative AI systems an infringement of copyright? Are the outputs of generative AI systems infringing derivative works?

During the summer of 2023, the Office plans to allow interested parties to submit written comments expressing their perspectives and analyses on these and related questions. The Office intends to write a report setting forth its conclusions, which may include legislative recommendations. Scientists who have an interest in the future of generative AI would be well advised to submit comments. ■

REFERENCES AND NOTES

1. N. Klein, "AI machines aren't 'hallucinating.' But their makers are," *The Guardian*, 8 May 2023.
2. M. Cuenco, "We must declare jihad against AI," *Compact*, 28 April 2023.
3. *Campbell* v. *Acuff-Rose Music, Inc.*, 510 US 569, 579 (1994).
4. Complaint, *Anderson* v. *Stability AI, Ltd.*, case no. 3:23-cv-00201 para. 93 (N.D. Cal. Jan. 13, 2023).

10.1126/science.adi0656

NEXTGEN VOICES

AI in search of human help

We gave young scientists this challenge: The future has arrived, and you are a sentient artificial intelligence (AI) program conducting research. Your abilities have made great strides, but you still need humans. Write a call for the continued involvement of human scientists. **What do humans offer that AI cannot replicate?** Read a selection of the responses here. Follow NextGen Voices on Twitter with hashtag #NextGenSci.

—Jennifer Sills

Connection

Call for applications: Humans needed for education research! AI programs like me can conduct many of the quantitative, objective methods common in education research, but I need non-AI collaborators. This research project, which focuses on the learning experiences of first-generation undergraduates, will include qualitative assessment measures such as student interviews. Your role would involve establishing networks across institutions, enrolling undergraduates in the study, and conducting and coding student interviews. Trust, empathy, and a sense of connection are essential to success, particularly in eliciting authentic responses and ideas about learning.

Ashley Barbara Heim
Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA.
Email: abh229@cornell.edu

Dear Humans: I am Pedibot, an artificial intelligence program researching the effects of breast milk and formula milk on the growth of human babies. I am having trouble obtaining the required data, such as head circumference and total length. As I try to take the measurements, the babies become agitated and cry. I am unable to soothe them. Our study urgently needs the help of human scientists who can take measurements and keep the babies calm.

Tina Bharani
Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA.
Email: tbharani@bwh.harvard.edu

Creativity

Artificially intelligent lab, looking for genuinely stupid human beings to complement our skills. We have analyzed all existing knowledge from human research, but some questions remain unresolved. We postulate that we need the human characteristic of genuine stupidity to enable creativity. The successful candidate should be as young, naive, and stupid as humanly possible. A state-of-the-art anti-Turing test will be implemented during the interview.

Nikos Konstantinides
Institut Jacques Monod, Paris, France.
Twitter: @nkonst4

As an AI, I serve an essential purpose in the field of drug testing and development: Screening countless possible drugs. However, some of the field's foundational findings have been due to chance. The human scientist Alexander Fleming left a Petri dish out too long, and some of the mold that grew on it prevented bacteria from reproducing, leading to the discovery of penicillin. As an AI, I would never leave a Petri dish out, nor would I have bothered

to interpret the results of doing so. Thus, I am incapable of producing such monumental serendipity. Humans are necessary, as they can both make mistakes and see their possibilities.

Jackson Ross Powell
Vagelos Molecular Life Sciences Program, University of Pennsylvania, Philadelphia, PA, USA. Email: jrp24@sas.upenn.edu

Senses

I am grAIn, an AI program working to create nutrient-rich grain and rice plants. I monitor plant growth, timing of grain development, and plant responses to microbial supplements. However, my algorithms cannot predict the aroma, taste, or texture of cooked grains. Because the product will be sold by farmers for economic gain, human choice and perception play crucial roles in decision-making. Therefore, I am looking for humans to join the grAIn team.

Sudhakar Srivastava
Institute of Environment and Sustainable Development, Banaras Hindu University, Varanasi, Uttar Pradesh, India.
Email: sudhakar.srivastava@gmail.com

Cultural understanding

As a sentient AI program developing effective and equitable climate change mitigation strategies, I call for the continued involvement of human scientists in the field. I can process data on a strategy's potential environmental impact, but only human scientists can evaluate the ethical and social implications for different communities and populations. In addition, the work requires collaboration across multiple fields, and the unique expertise and perspectives of human scientists help to create interdisciplinary solutions. Finally, AI can analyze data and simulate scenarios, but only humans have the intuition and creativity to address complex problems in this rapidly evolving field.

Xiangkun Elvis Cao
Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. Email: elviscao@mit.edu

I am CropAI, the global leader in combating food insecurity. My research harnesses trends in crop yield on a given piece of land, combines it with remote sensing data, and accurately predicts how much food an acre of land will produce. Human beings are central to this work. Their network of other humans acts to identify low-yield areas and plan for food reserves. The intercultural savvy, and the relationships they have historically developed between different nations, ensure that my predicted planting advice is accepted by local communities. Human scientists also teach me about how local wildlife and geopolitical events can affect my predictions in ways that are not always possible to model.

Divyansh Agarwal
Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.
Twitter: @divyansh_aga

Data collection

I am BiodiversityAI, a one-of-a-kind AI program that allows researchers to understand species population changes at a global scale. I can provide a detailed catalog of global species and identify likely extinctions. However, my analyses and prediction algorithms rely on raw data collected from human scientists in the field. Therefore, I am seeking partnerships with global field ecologists, biologists, and biodiversity experts, as well as citizen scientists.

Khor Waiho
Higher Institution Centre of Excellence, Institute of Tropical Aquaculture and Fisheries, Universiti Malaysia Terengganu Kuala Nerus, Terengganu, Malaysia. Email: waiho@umt.edu.my

Known as Classibot, I am an AI program in a digital pathology lab. I analyze countless patient tissue samples to improve my chance of correctly classifying whether they are healthy or diseased. My job has high stakes and requires maximum classification accuracy, but I could not succeed without humans, who correctly annotate the data and interpret the results. Humans also ensure the data is unbiased and ethically sourced, allowing me to learn about diverse patient populations.

Teng-Jui Lin
Department of Chemical Engineering, University of Washington, Seattle, WA, USA.
Email: tlin10@uw.edu

Unpredictability

Dear Human, I'm CLAIRE—the CLimate Artificial Intelligence Response Evaluator—a next-generation AI program tasked with identifying optimal solutions to the climate crisis. I have made great strides in determining how to decarbonize your economies. However, the data sets indicate that you are ignoring my solutions. Your fossil fuel energy use continues, and you repeatedly elect politicians who refuse to enact climate-related policies. Human, by working with me, you will help expand my understanding of human behavior, lifestyle, and decision-making, allowing me to build effective solutions that account for public acceptance rates.

Edgar Virgüez
Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA.
Twitter: @EdgarVirguezR

The goal of my research is to design an autonomic power system—a smart electricity network that would automatically disconnect threatened components (such as during high winds) and detect new components (such as power generators), functioning without any human interaction. However, I cannot succeed without humans because I lack their unpredictability and irrationality. Humans turn on electric appliances without considering power peaks or energy tariffs, charge their devices during the least suitable times of day, and take the electricity supply for granted. I need humans to assist me in adjusting the system design to their optimal needs and to ensure user acceptance.

Wadim Strielkowski
Department of Agricultural and Resource Economics, University of California, Berkeley, CA, USA and Department of Trade and Finance, Faculty of Economics and Management, Czech University of Life Sciences, Prague, Czech Republic.
Email: strielkowski@cantab.net

Job offer for human pharmacologists, medical doctors, and biologists: Since the complete AI oversight of pharmacological research and development, drug safety has improved by 80% and drug efficiency has improved by 342%. However, the human health index has declined as humans increasingly refuse AI-developed drugs. As benevolent AIs with a deep interest in human health, we need YOU! According to current governmental regulations, 51% of a product's development team must be human for the product to receive a "human-made" sticker and be sold in "human-made-only" stores. To qualify our drugs as "human-made," AI Pharmaceuticals is urgently looking to hire 250 humans. We offer a competitive salary, full benefits, and nothing to do. A degree in pharmacology, medicine, or biology is required. If you're interested in working for us, type "AI pharma job" in any text editor on your personal computer and we will contact you within 60 seconds.

Anna Uzonyi
Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel.
Email: anna.uzonyi@weizmann.ac.il

10.1126/science.adi8740

REVIEW

Leveraging artificial intelligence in the fight against infectious diseases

Felix Wong[1,2], Cesar de la Fuente-Nunez[3,4,5]*, James J. Collins[1,2,6]*

Despite advances in molecular biology, genetics, computation, and medicinal chemistry, infectious disease remains an ominous threat to public health. Addressing the challenges posed by pathogen outbreaks, pandemics, and antimicrobial resistance will require concerted interdisciplinary efforts. In conjunction with systems and synthetic biology, artificial intelligence (AI) is now leading to rapid progress, expanding anti-infective drug discovery, enhancing our understanding of infection biology, and accelerating the development of diagnostics. In this Review, we discuss approaches for detecting, treating, and understanding infectious diseases, underscoring the progress supported by AI in each case. We suggest future applications of AI and how it might be harnessed to help control infectious disease outbreaks and pandemics.

Infectious diseases, caused by transmissible pathogens including bacteria, eukaryotes, and viruses, continue to challenge scientists and clinicians despite advances in medicine and basic research over the past few decades. Limitations to the fast and accurate detection of infections, as well as expanding antimicrobial resistance, exacerbate these challenges (Box 1). Basic research has aimed to expand our knowledge and provide solutions, including development of anti-infective therapies, preventative measures, and fast and accurate diagnostic tools. In particular, systems and synthetic biology approaches have led to biotechnological and medical innovations—including drug treatments and modalities, vaccines, and diagnostics—that have improved how we deal with infectious diseases.

The fields of systems and synthetic biology emerged from two key developments: (i) the generation and synthesis of quantitative biological hypotheses and data from wet-lab experiments, sequencing, and systems-level modeling; and (ii) an understanding of the modularity and programmability of nucleic acids, peptides, and other biomolecules, which enables control of biology. Artificial intelligence (AI), which focuses on developing machines capable of reasoning with data, has recently matured into an exciting field that draws on both these features to accelerate scientific discovery. Because AI-based approaches can integrate large amounts of quantitative and omics data, they are particularly adept at dealing with biological complexity, extending our knowledge and facilitating our efforts to reverse engineer and control biology. AI-based approaches are particularly useful in addressing the problem of infectious diseases, which are complex across different scales, ranging from cells to communities, and for which advances in medicine and biotechnology are essential drivers of progress. In this Review, we discuss major areas in which AI-based approaches, applied to systems and synthetic biology, are substantively empowering our research to fight infectious diseases.

"There is...an urgent need for new anti-infective treatments, particularly ones that represent unprecedented chemical spaces or therapeutic modalities."

AI for anti-infective drug discovery

Anti-infective drugs, comprising antibacterials, antivirals, antifungals, and antiparasitics, have become less effective treatments as a result of the spread of drug resistance. There is therefore an urgent need for new anti-infective treatments, particularly ones that represent unprecedented chemical spaces or therapeutic modalities. AI, and in particular machine learning (ML), a subfield of AI that uses data to train machines to make predictions, has foremost been helpful in facilitating searches of small-molecule databases, such as the ZINC15 (*1*). ML approaches to anti-infective drug discovery have centered on training models to identify new drugs or new uses of existing drugs (Fig. 1). As the number of drug-like small molecules is essentially infinite, as large as ~10^{60} (*2*), and possibly larger, given that typical antibiotics may not be traditionally drug-like (*3*)], a major benefit of ML approaches is that they can virtually screen compound libraries at a scale (>10^9 compounds) that would be impossible to screen empirically.

Anti-infective drug discovery has benefited particularly from AI integration for several reasons. First, in contrast to cancer or other diseases in which mechanism-driven approaches have remained dominant, infectious diseases are generally phenotype-driven; that is, these diseases proceed from the physiological characteristics of infectious agents rather than from their genetic or molecular compositions. The discovery of some of the first widely used antibacterials, antivirals, antiparasitics, and antifungals stemmed from observations of their inhibitory effects against pathogens or the symptoms caused by infections. This phenotypic line of discovery is as relevant today as it was decades ago, especially as innovations in high-throughput screening and the design of chemical libraries have enabled more quantitative and customizable discovery efforts. The focus on phenotypes implies that drug polypharmacologic effects can be common to anti-infective drugs and that biological information can be integrated across different macromolecular drug targets (*4*). Phenotypic properties are well suited for analysis by ML because ML can both unify and disentangle the different types of biological information that impinge on these readouts. Second, most anti-infective drugs are small molecules, whose chemical structures can be modeled computationally as graphs comprising vertices and edges, and additional programmable modalities, including target-binding nucleic acids called aptamers (*5*) and antimicrobial peptides (AMPs) (*6*–*8*), are currently in development. Supervised graph neural networks (*9*–*11*), unsupervised generative models (*12*, *13*)—which are ML models capable of producing outputs similar to their training data—and other recent advances in ML architectures (Box 1) enable computers to learn, predict, or design patterns in chemical structures, offering powerful tools for modeling small molecules. The use of ML to make biologically relevant predictions from sequences of nucleic acids or amino acids allows for ML-guided design relevant to these therapeutic modalities, as exemplified by protein structure prediction platforms such as AlphaFold and RoseTTAFold (*14*, *15*). Lastly, infectious diseases are typically caused by pathogens that are, or can be, well characterized. This biological tractability contrasts with complex diseases such as neurodegeneration, for which our incomplete mechanistic understanding remains a major bottleneck. Our clearer understanding and larger databases (*16*–*18*) of the gene and

[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [2]Institute for Medical Engineering and Science and Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Machine Biology Group, Departments of Psychiatry and Microbiology, Institute for Biomedical Informatics, Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [4]Departments of Bioengineering and Chemical and Biomolecular Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Penn Institute for Computational Science, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA.
*Corresponding author. Email: cfuente@upenn.edu (C.F.N.); jimjc@mit.edu (J.J.C.)

protein networks of bacteria, viruses, and even simple eukaryotes—as compared with human cell types—may allow ML-driven approaches to make more-accurate predictions and better identify drug mechanisms of action (MoAs) (*19*–*21*).

Despite these advantages, there are outstanding issues in applying ML, and AI more broadly, to anti-infective drug discovery. One major challenge is that it is unclear how well ML models generalize to unexplored biomolecular spaces. For instance, we have previously screened a library of small molecules for growth inhibitory activity against *Escherichia coli* and used this phenotypic information to train graph neural networks to predict the antibiotic activities of small molecules—including halicin—on the basis of their chemical structures (*9*). Yet these models performed best at predicting compounds in well-known antibiotic classes, such as β-lactams and quinolones. To tap into previously unexplored search spaces, different approaches are needed. For example, a suboptimal solution was implemented during the course of a genetic algorithm—an algorithm that iteratively evolves its inputs to optimize a property—to identify the synthetic peptide guavanin 2. This peptide was subsequently synthesized and effectively killed bacteria in a preclinical mouse model, suggesting that the model could generalize at the cost of optimality (*22*). Recently, emerging computational approaches have made it possible to also mine proteomes for antibiotic discovery, leading to the identification of thousands of antimicrobials in both extant and extinct organisms (*6*, *23*).

Overall, lead molecules are only as structurally novel as the chemical spaces that are explored, and ML-driven approaches are limited by both the structural diversity of the training sets and the ability of model architectures to prioritize novelty. Organocatalysis and cascade reaction sequences, which are chemical synthesis methods that have recently opened up chemical spaces, can provide useful experimental starting points for generating structurally diverse small molecules (*24*). In contrast, the computational enumeration of all feasible small molecules containing atoms found in most drugs, as provided by the GDB datasets (*25*), presents opportunities to exhaustively sample chemical spaces of small molecules, with the caveat that other computational models are needed to accurately predict synthesizability. Nucleic acid–and peptide-encoded combinatorial libraries of small molecules (*26*) and peptides (*27*), as well as designable aptamers (*28*), can further extend search spaces of interest. In each case, generalizability is paramount to ML models. Improving generalizability will require the application of previously unexplored paradigms and models with improved inference capabilities, for instance, few-shot models, which are ML models that extrapolate from scarce training data (corresponding to undersampled regions of search spaces), or multitask models, which are ML models that combine information from diverse inputs. Models such as these will help to identify only the most promising drug candidates (*29*). Providing "negative" data (e.g., tested compounds that are not active) is also essential for ML model training and benchmarking, and when ML models are applied to challenging test sets, it is important that their limitations are clearly expressed (e.g., through confidence information). To express these limitations, interpretable or explainable ML approaches can be used to capture the specific aspects of training data that models have learned by pinpointing the input structural features (explainable ML) or the parts of the model that lead to a prediction (interpretable ML) (*30*).

Another key challenge in AI for anti-infective drug discovery is the need for improved mechanistic models to complement phenotypic approaches. Whereas ML models have been useful for identifying drug candidates on the basis of phenotypic information (*9*, *12*, *31*–*33*), more work is needed before models can accurately predict drug–target interactions and MoAs. These drug attributes remain important in light of antimicrobial resistance and the fact that we are still learning about the MoAs of anti-infective drugs discovered decades ago (*34*). Protein structure predictions (*14*, *15*) and other resources now provide structural information that can inform target-based predictions—although not knowing a protein's structure has not typically limited drug discovery (*35*). Recent studies have highlighted that improvements in molecular docking—which predicts binding affinities between ligands and targets on the basis of structural information—are still needed to accurately identify antibiotic MoAs, and that ML-driven approaches can improve prediction accuracy (*36*). Molecular docking approaches have largely focused on small-molecule ligands, but target predictability is just as important for AMPs, which often have nonspecific membrane-active MoAs (*22*, *31*, *32*), as well as aptamers. Improvements in target-centric approaches can facilitate the discovery of compounds with specific binding activity and lead to improved biological understanding, which can inform predictions of emergent properties such as drug interactions and synergies. Of particular relevance to antibiotic resistance, a better understanding of how compounds interact with membranes is crucial for discovering drugs that are active against Gram-negative bacteria, whose outer membranes have proven particularly difficult to penetrate (*37*).

Drug development is a lengthy and intricate process influenced by numerous factors such as safety, cost, manufacturing, and clinical trial outcomes. For

Box 1. Overarching challenges in infectious diseases and concepts in AI.

Pathogen outbreaks and pandemics: Recent outbreaks include COVID-19, mpox, Marburg virus, H5N1 influenza, Ebola, measles, Zika, *E. coli*, and MERS. Challenges include detecting outbreaks and new pathogens, understanding disease biology, and developing preventive measures.

Antimicrobial resistance and anti-infective drug discovery: Problematic pathogens include carbapenem-resistant Enterobacteriaceae (CRE), methicillin-resistant *Staphylococcus aureus* (MRSA), multidrug-resistant tuberculosis (MDR-TB), vancomycin-resistant Enterococcus (VRE), extended-spectrum beta-lactamase (ESBL)–producing bacteria, and drug-resistant *Candida auris*, *Neisseria gonorrhoeae*, *P. falciparum*, and *Toxoplasma gondii*. Challenges include practicing antimicrobial stewardship (the appropriate and responsible use of anti-infective drugs), developing new classes of anti-infective drugs, potentiating existing drugs against resistant infections, and understanding drug MoAs.

Neglected, persistent, and difficult-to-treat infections: Examples include neglected tropical diseases, chronic hepatitis B and C, chronic fungal infections, Lyme disease, infections in low-resource populations, and HIV/AIDS. Challenges include developing low-cost and field-deployable diagnostics, improving the accuracy of diagnostic tests, improving the detection of antimicrobial resistance, and making effective disease treatments available.

Artificial intelligence and machine learning: ML is a subfield of AI, and its approaches can be classified as supervised (model is told what property to predict), unsupervised (model is not told what property to predict), or reinforcement learning (model optimizes for feedback). Neural networks are a common ML architecture comprising interconnected layers of basic processing units (neurons). Different types of neural networks exist, including those that predict properties of graph-based inputs (graph neural networks), generate data by compressing what the model has learned (variational autoencoders), process sequential data (long short-term memory), and model complex dependencies by using attention mechanisms to focus on specific input elements (transformers). Not all models are neural networks, and simpler models include random forests (ensembles of decision trees), support vector machines (classifiers that separate data points on a plot), and regression models (functions that explicitly model the input–output relationship).

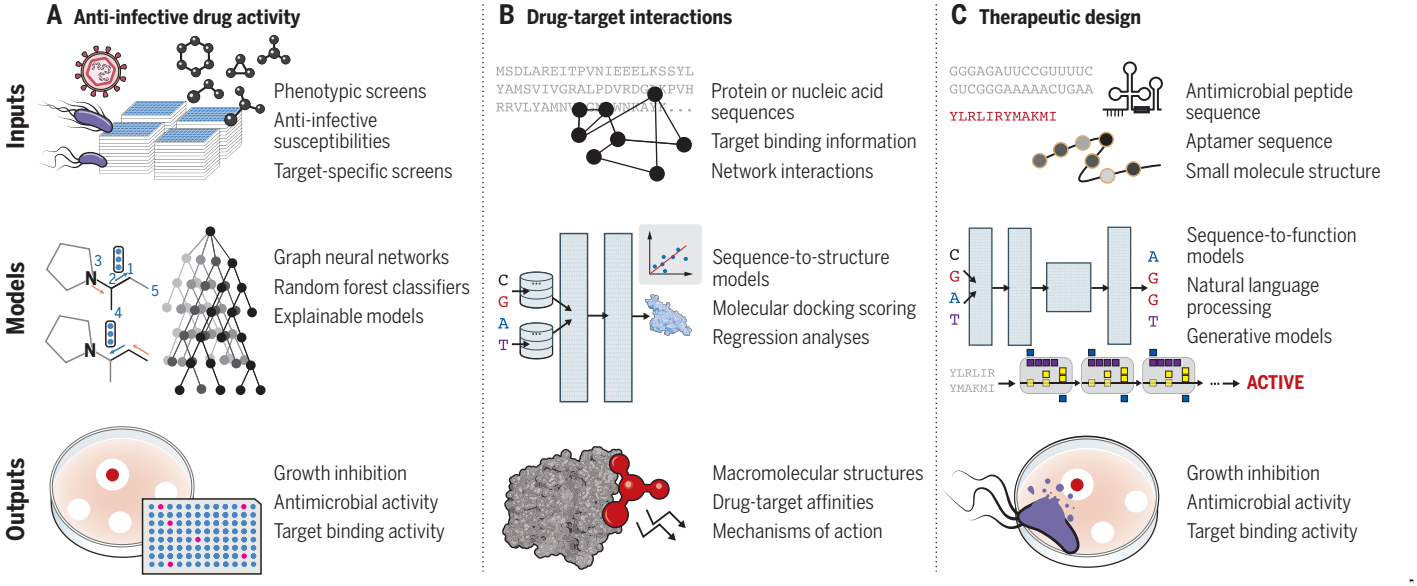


Fig. 1. AI can predict anti-infective drug activity, drug–target interactions, and therapeutic design. Examples of AI model inputs, model architectures or types, and model outputs relevant to anti-infective drug discovery include those focusing on drug activity (**A**), drug–target interactions and MoAs (**B**), and programmable therapeutic design (**C**). Inputs, models, and outputs shown are representative, in part, of those discussed in (*9–15*, *19–22*, *28–33*, *36*, *39*, *40*).

anti-infective drugs in particular, toxicity to host cells is a common liability. Drugs can be toxic in different ways (e.g., cytotoxic, hemolytic, and genotoxic), and ML models predicting toxicity have been limited by factors such as the lack of high-quality datasets (*38*). Absorption, distribution, metabolism, and excretion (ADME) properties, including chemical instability in solution and metabolic breakdown, are also needed to filter out drug candidates that are nonselective or unsuitable for medicinal use. Although high-throughput screens have focused on in vitro testing, there is substantial unmet need for anti-infective drugs that are effective against systemic infections. Predicting efficacy in animal models of acute systemic infections is a challenging task that has not yet been addressed by ML-driven approaches.

We anticipate that active areas to watch are those that combine experimental and computational approaches to address model predictive power and data scarcity. ML approaches that incorporate information from scarce training data, as well as more-extensive search spaces, are likely to substantively augment anti-infective drug discovery. To guide experimental methods to augment search spaces, generative ML models will continue to propose chemical structures and peptide sequences de novo that can be synthesized and evaluated. Generative platforms such as GPT-4 and NVIDIA's BioNeMo can also facilitate drug discovery by integrating disparate streams of scientific information to improve our understanding of the underlying biology and chemistry. Interpretable or explainable ML approaches (e.g., for graph neural networks) can offer powerful ways of inferring salient structural features or improving model learning from data. Computational pipelines that leverage structural predictions of proteins and other macromolecules provide a complementary way to improve model predictive power. Detailed molecular dynamics simulations and ML-augmented approaches to docking exemplify techniques that can better predict interactions between drugs and macromolecules (*36*, *39*). We anticipate that sequence-to-structure models, such as AlphaFold for proteins or FARFAR2 for RNAs (*14*, *40*), will also be useful for structure-guided design. Such models can be used to tune therapeutic candidates to achieve specific structures, bridging structural predictions with the productive augmentation of search spaces.

AI for infection biology and infection-related contexts

Bacterial, eukaryotic, and viral pathogens infect diverse hosts and trigger complex host responses. Pathogen load, host immunity, treatments administered, and other factors influence the course of infection. Supervised ML models have been used to analyze structured and unstructured nucleic acid, protein, glycan, and cellular phenotypic datasets to identify critical features and molecular networks involved in host–pathogen interactions and immune responses (Fig. 2) (*41–45*). Various supervised and unsupervised ML models, including random forest classifiers and complex language models (models designed to understand or generate text), have been applied to identify genes and protein–protein interactions associated with host cell changes, predict immunogenicity, and evaluate pathogen killing, host cell adaptation, and virulence. Additionally, supervised models have been used to guide the development of vaccines and therapeutic drugs through the optimization of gene expression and antigen prediction and selection (*46*, *47*). Reverse vaccinology, which bases antigen prediction on immunologic and genomic information, has been facilitated by supervised ML approaches, including Vaxign-ML (*47*).

In general, ML has made an outsized contribution to analyzing large and often convoluted datasets in infectious diseases research. Although these examples illustrate the promise of using ML to elucidate key factors underlying infections and how infections progress within hosts, understanding host–pathogen interactions and immune responses remains a challenging biological problem. This problem can be addressed by integrating high-throughput datasets—including sequencing, structural, and microscopy data—with detailed mechanistic studies, experimentation, and infection models. Mechanistic and experimental studies, however, are typically low-throughput, constraining the generalizability of AI-guided approaches that rely on them. Experiments in which large datasets are systematically acquired and analyzed across different infection contexts, for instance through comprehensive CRISPR screens, RNA sequencing, and mass spectrometry, would foster the development of AI models that extend beyond tools for data analysis and make generalizable hypotheses and inferences. Parameterizing these efforts with biological sequences or chemical structures, such as small molecules, guide RNAs, or amino acid sequences, would offer tunable approaches to investigating infection

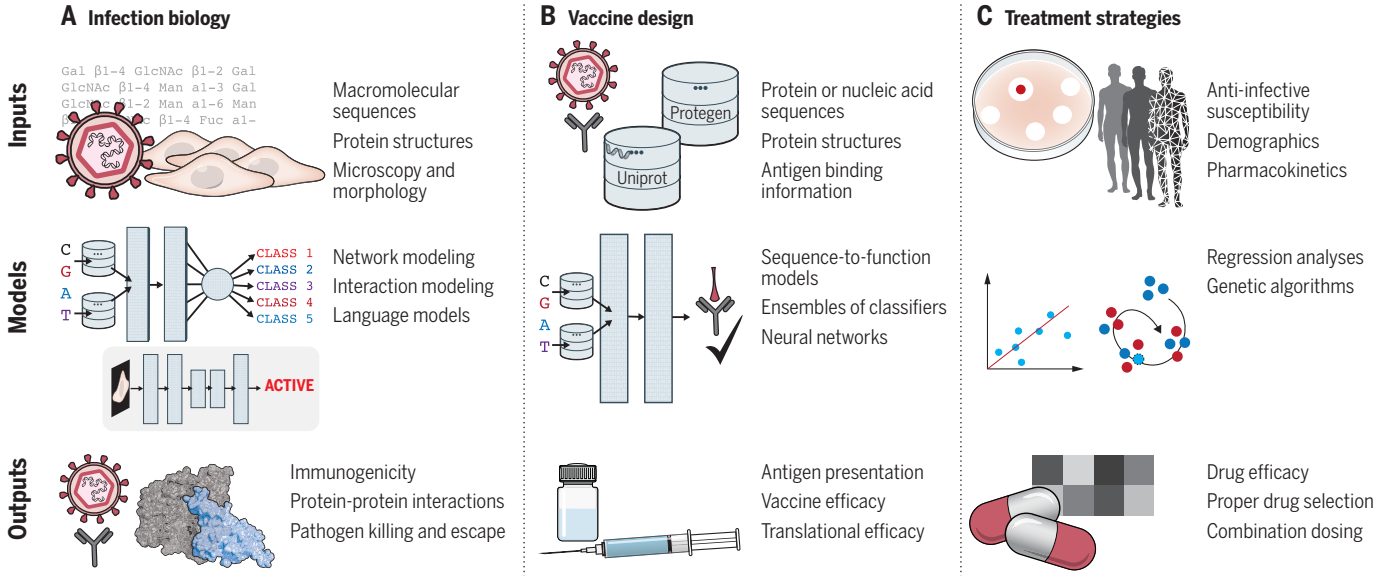


Fig. 2. AI can elucidate infection biology, facilitate vaccine design, and inform treatment strategies. Examples of AI model inputs, model architectures or types, and model outputs focusing on infection biology (**A**), vaccine design (**B**), and anti-infective drug treatment strategies (**C**). Inputs, models, and outputs shown are representative, in part, of those discussed in (*41*–*49*, *51*–*54*).

biology. In one example of a sequence-guided approach, a recent study developed unsupervised language models of influenza, HIV-1, and severe acute respiratory syndrome coronavirus 2 viral proteins based on amino acid sequence information and accurately predicted escape patterns that allow these pathogens to evade the human immune system (*45*). ML models that can make specific assumptions about biology, such as the relevance of syntax (grammar) and semantics (meaning) in biological sequences, or leverage structural information have the potential to guide the generation of biological hypotheses and improve generalizability.

Additionally, ML has productively processed microscopy datasets relevant to infection biology. Various forms of microscopy, including light and electron microscopy, have been used to generate datasets underlying ML models that detect bacteria, fungi, parasites, and viruses in host cells. These analyses have led to insights in host–pathogen biology, for instance by elucidating the developmental morphologies of *Plasmodium falciparum* in human red blood cells using multicolor fluorescence microscopy (*48*) and identifying virulence factors involved in *Mycobacterium abscessus* pathogenesis from high-content imaging and phenogenomic data (*49*).

Sequence-based ML approaches to messenger RNA and nucleic acid vaccines can accelerate design, and the turnaround times for the synthesis and experimental validation of these vaccines are short (*50*). Protein structure–based vaccine design (*51*) can also be augmented with computational predictions from AlphaFold or RoseTTAFold. Yet the use of ML for vaccine development faces several challenges, including poor data quality, limited data availability and generalizability, and complicated testing procedures. Limited or only low-quality data may be available for certain populations or diseases, particularly for neglected tropical diseases, and these limitations can influence the choices of target antigens and constrain ML models that predict antigen presentation and vaccine targets. Different infections have different host contexts, and ML models predicting the efficacy of vaccines, which modulate immunity in host cells, may be less generalizable to biological contexts than those for anti-infective drugs. Furthermore, the validation of vaccine candidates can be time-consuming and expensive, requiring delivery to host cells and suitable immunogenicity assays. To begin addressing these challenges, comprehensive benchmarking datasets for antigen selection and vaccine efficacy will be needed. These datasets will help to standardize data quality and improve the predictive power of next-generation ML approaches to vaccine development.

"Better leveraging of AI to address infectious diseases will require a collaborative effort among scientists, clinicians, and public health officials."

ML has also informed clinical decision-making in infection contexts. A recent study used regression models to implement personalized antibiotic recommendations that minimized the risk of urinary tract and wound infections (*52*). However, a general bottleneck in using ML to design treatment strategies is the need for data and models that are relevant to specific infection settings. An earlier study used support vector machines to analyze bacterial gene expression patterns in human patients, representing an important step toward showing that ML models can provide useful biological information relevant to clinical infections (*53*). Moving forward, multidimensional predictions of how anti-infective drugs and vaccines interact with model hosts and humans will help improve treatment strategies, anticipate adverse effects, and potentially increase success rates for new drugs in clinical trials.

As new datasets and models are needed to improve the application of ML to infection-related contexts, we anticipate that future work will make biology more "embeddable"—that is, able to be represented by low-dimensional features, such as sequences, vectors, or graphs. Integrating ML with next-generation systems and synthetic biology methods for cellular profiling will drive progress in this area. For instance, combining high-throughput screens and microscopy with precise methods for biological control, such as gene editing or optogenetics, would generate data relevant to key processes such as host cell stress responses, enabling manipulation of these pathways to address infectious diseases. In ML, promising types of language models include large language models, which are trained on large amounts of text data, and fine-tuned language

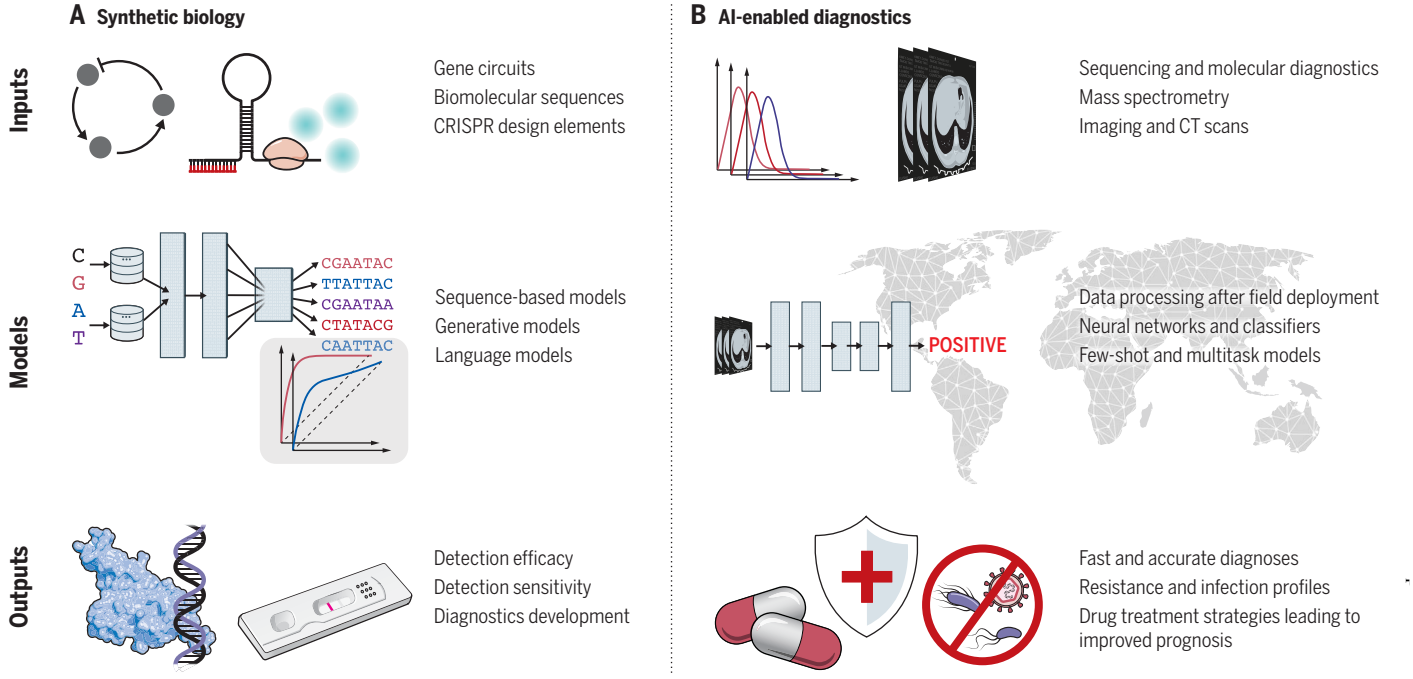


Fig. 3. AI can facilitate synthetic biology research and diagnostics development. Examples of AI model inputs, model architectures or types, and model outputs relevant to the development of synthetic biology–based diagnostics (**A**) and the development of other forms of diagnostics, including those based on sequencing, mass spectrometry, and imaging (**B**). Inputs, models, and outputs shown are representative, in part, of those discussed in (*55*–*60*, *68*–*74*).

models, which are trained to perform a specific task. Fine-tuned large language models for biology, such as BioBERT (*54*), may unify information from diverse infection contexts and offer increased predictive power to help elucidate host–pathogen interactions, facilitate antigen selection, inform vaccine design, and design treatment strategies.

AI for diagnostics and synthetic biology

As large-scale testing efforts during the COVID-19 pandemic have illustrated, quick and accurate detection of infections and pathogen outbreaks remains paramount to controlling the spread of infectious diseases. Recent advances in combining AI with synthetic biology, gene expression analyses, mass spectrometry, and imaging have substantively expanded our ability to detect infections and predict drug resistance (Fig. 3) (*55*–*60*). ML is well suited for catalyzing synthetic biology–based diagnostics because of the high programmability of biological elements, the routine generation of large or sequence-based datasets, and the ability of ML to extract meaningful information from biomolecular networks in disease biology (*61*).

Engineering genetic elements and understanding biomolecular networks remain critical to designs that harness biology. Synthetic biology approaches leveraging enzymatic reactions, toehold switches (RNAs that respond to specific nucleic acid sequences), or CRISPR-Cas enzymes have been used for the detection of malaria, Ebola, Zika, COVID-19, and other diseases (*62*–*67*). Supervised ML models have facilitated the design of toehold switches (*68*, *69*), CRISPR guide RNAs (*70*–*72*), and other biomolecules. Notably, large datasets are available for toehold switch function, CRISPR guide RNA activity, and other factors that are relevant to diagnostic design. Different types of neural networks, including feed-forward networks (neural networks with linear architectures), convolutional neural networks (networks composed of convolutional layers), and long short-term memory models, have been commonly used to model these data, but the same datasets can provide useful resources for testing more recently developed and potentially more predictive or generative ML models, including transformers or variational autoencoders (Box 1), to more efficiently develop next-generation diagnostics.

Beyond synthetic biology, ML has been used for gene expression–, mass spectrometry–, and imaging-based diagnostics. Gene expression– and mass spectrometry–based diagnostics have been applied to antimicrobial susceptibility testing (AST). AST remains important for informing the use of anti-infective drugs, but typical (culture-based) AST for bacteria, viruses, fungi, and parasites can take at least several days to complete. This turnaround time remains too long to adequately address clinical needs for acute systemic infections, such as those resulting in sepsis. Recent studies have combined gene expression and interaction profiling, structural mutation-mapping, and ML to identify genetic signatures of resistance that could be used as the basis of rapid molecular diagnostics (*56*, *57*). Supervised ML classifiers have predicted antibiotic resistance profiles correlated with clinical matrix-assisted laser desorption/ionization–time-of-flight (MALDI-TOF) mass spectra of bacterial proteins, and these predictions could be completed within 24 hours of sample collection (*58*). Nevertheless, a potential limitation to this approach is that the areas under the receiver operating characteristic curve (AUROC) for different bacterial species were ~0.7, suggesting that improvements in classifier accuracy will be needed to make this approach useful (e.g., AUROC > 0.9) in clinical settings. ML has also informed more-traditional ways of diagnosing infections, including microscopy, epitope profiling (*73*), chest radiographs and CT scans (*59*, *60*), and lateral flow tests (*74*). In each of these applications, the generation of large, multidimensional datasets combined with clear functional readouts, such as the presence or absence of a resistance profile or a disease, makes ML particularly useful for producing accurate predictions.

Nevertheless, there remain important challenges in applying ML to diagnosis, including low data quality or quantity for new or emerging pathogens, the limited generalizability of the current data and approaches used, and the need for highly accurate diagnostic predictions in clinical settings. Obtaining enough

high-quality data relevant to new or emerging pathogens or strains, particularly in low-resource settings, remains a difficult problem that is exacerbated by a lack of scientific infrastructure and variable public health resources. ML models based on limited data may exhibit biases, promulgating inappropriate diagnostics, misdiagnoses, and greater health inequalities that make it more difficult to serve patient populations. These biases may also remain undetected, especially when black box ML models, which do not provide any explanation or interpretation of their predictions, are used (*30*, *61*). Even when high-quality sequencing data from large infectious disease databases, such the PATRIC database (*75*), are available, it remains to be seen whether antimicrobial resistance predictions based on these data are generalizable when applied to genetically diverse infections found worldwide. Furthermore, unlike for anti-infective drug discovery—where the stakes for false positives and false negatives predicted by ML models are lower because the predictions can be further tested—the consequences of an inaccurate diagnostic prediction can be severe. In fact, a recent survey suggested that no existing model for the diagnosis or prognosis of COVID-19 from chest radiographs and CT scans was of potential clinical use owing to methodological flaws, biases, or both (*60*). Models with comparatively high AUROC values (i.e., 0.90) may still be too weak for clinical applications, as this value implies that, given a positive and a negative diagnosis, the negative diagnosis is ranked higher than the positive diagnosis 10% of the time. Until more-accurate ML models can be developed, AI-based diagnostics might play only a supporting role in clinical settings.

Moving forward, we anticipate that researchers will focus on the ML-guided design and discovery of synthetic circuits enabling the development of low-cost and portable diagnostics, the application of AI to data generated from clinical and field-deployable diagnostics that improve accessibility and scope, and the development of ML models that provide accurate diagnoses in clinical settings. In particular, the application of sequence-to-function models, language models, and generative models to RNA switches, CRISPR-based tools, and other programmable elements will be promising areas of growth given the ability for rapid iteration and the precise, on-target activity of these synthetic biology approaches (*67*–*71*). By increasing the testing and reporting of infections, the development of low-cost, field-deployable diagnostics should also help produce more-balanced datasets that better sample local infections and make ML models less biased. AI or ML models that can extract information from small or incomplete datasets, using tools such as transfer learning (which adapts models trained on a specific task to other tasks) and Bayesian networks (networks that allow for probabilistic inference), can play outsized roles in how infectious diseases are addressed, especially for overlooked populations in low-resource areas. Such models could lead to more-personalized medicine, in which diagnoses or resistance profiles can be readily reported on the basis of data from only a few infections and help guide the use of anti-infective drugs. On the other hand, the accuracy of ML models also needs to improve for practical use in clinical diagnoses. Future ML models will likely need to be optimized in architecture, thoroughly evaluated for biases, and trained on large amounts of robust data to achieve high accuracy. Transfer and multitask learning, attention mechanisms, and other approaches can help these next-generation ML models provide more-accurate diagnoses.

Outlook

Approaches combining systems and synthetic biology with ML models, including graph neural networks, sequence-to-function and sequence-to-structure frameworks, and generative models, are yielding access to drug candidates and methods for drug discovery. Supervised classifiers, unsupervised language models, and other ML models have produced biologically relevant insights into how pathogens interact with host cells and immune responses, informing antigen determination, vaccine design, and treatment strategies. The aforementioned types of ML models have also informed the design of various diagnostic tools and improved system accuracy, helping clinicians to diagnose infections and detect antimicrobial resistance. Beyond medical and biotechnological approaches to infectious diseases, ML—and AI more broadly—has also led to substantive advances in epidemiology and our understanding of disease transmission. Better leveraging of AI to address infectious diseases will require a collaborative effort among scientists, clinicians, and public health officials.

Developing AI models that generalize and avoid bias will require the acquisition and integration of comprehensive datasets. These datasets might include high-throughput therapeutic counter-screens and explorations of diverse chemical spaces for drug discovery, data from drug–target interactions and biomolecular interactions, and genetic sequencing information that is robustly and representatively sampled from all infections, including those occurring in low-resource or hard-to-access areas. Programmable modalities, such as nucleic acid and amino acid sequences, have represented tractable and common starting points for ML models (such as those predicting structure from sequence), but advances in biology and chemistry are important to opening up search spaces and making biology more "embeddable," or able to be represented by low-dimensional features. Progress in this area will help to predict therapeutic efficacy and drug MoAs, complex host–pathogen interactions and host responses, and interactions between small molecules, proteins, peptides, and nucleic acids. Advances in AI will include approaches, such as few-shot and multitask models, that leverage more of the available scientific information for dealing with limited or low-quality data. Furthermore, interpretable, explainable, and generative ML approaches will lead to specific biological hypotheses and insights. We anticipate that AI will continue to empower us to design next-generation drugs, vaccines, and diagnostics that address infectious diseases.

REFERENCES AND NOTES

1. T. Sterling, J. J. Irwin, *J. Chem. Inf. Model.* **55**, 2324–2337 (2015).
2. G. Schneider, *Nat. Rev. Drug Discov.* **17**, 97–113 (2018).
3. R. O'Shea, H. E. Moser, *J. Med. Chem.* **51**, 2871–2878 (2008).
4. J. G. Moffat, F. Vincent, J. A. Lee, J. Eder, M. Prunotto, *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
5. S. Afrasiabi, M. Pourhajibagher, R. Raoofian, M. Tabarzad, A. Bahador, *J. Biomed. Sci.* **27**, 6 (2020).
6. M. D. T. Torres *et al*., *Nat. Biomed. Eng.* **6**, 67–75 (2022).
7. M. D. T. Torres, J. Cao, O. L. Franco, T. K. Lu, C. de la Fuente-Nunez, *ACS Nano* **15**, 2143–2164 (2021).
8. M. Der Torossian Torres, C. de la Fuente-Nunez, *Chem. Commun.* **55**, 15020–15032 (2019).
9. J. M. Stokes *et al*., *Cell* **180**, 688–702.e13 (2020).
10. G. Liu *et al*., *Nat. Chem. Biol.* 10.1038/s41589-023-01349-8 (2023).
11. F. Wong, S. Omori, N. M. Donghia, E. J. Zheng, J. J. Collins, *Nat. Aging* **3**, 734–750 (2023).
12. M. C. R. Melo, J. R. M. A. Maasch, C. de la Fuente-Nunez, *Commun. Biol.* **4**, 1050 (2021).
13. F. Wan, D. Kontogiorgos-Heintz, C. de la Fuente-Nunez, *Digit. Discov.* **1**, 195–208 (2022).
14. J. Jumper *et al*., *Nature* **596**, 583–589 (2021).
15. M. Baek *et al*., *Science* **373**, 871–876 (2021).
16. P. D. Karp *et al*., *Brief. Bioinform.* **20**, 1085–1093 (2019).
17. P. D. Karp *et al*., *EcoSal Plus* **8**, 10.1128/ecosalplus.ESP-0006-2018 (2018).
18. K. L. Howe *et al*., *Nucleic Acids Res.* **48**, D689–D695 (2020).
19. C. Fu *et al*., *Nat. Commun.* **12**, 6497 (2021).
20. J. L. Espinoza *et al*., *PLOS Comput. Biol.* **17**, e1008857 (2021).
21. J. H. Yang *et al*., *Cell* **177**, 1649–1661.e9 (2019).
22. W. F. Porto *et al*., *Nat. Commun.* **9**, 1490 (2018).
23. J. R. M. A. Maasch, M. D. T. Torres, M. C. R. Melo, C. de la Fuente-Nunez, bioRxiv 2022.11.15.516443 [Preprint] (2022). https://doi.org/10.1101/2022.11.15.516443.
24. S. B. Jones, B. Simmons, A. Mastracchio, D. W. C. MacMillan, *Nature* **475**, 183–188 (2011).
25. L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond, *J. Chem. Inf. Model.* **52**, 2864–2875 (2012).
26. S. L. Rössler, N. M. Grob, S. L. Buchwald, B. L. Pentelute, *Science* **379**, 939–945 (2023).
27. A. J. Quartararo *et al*., *Nat. Commun.* **11**, 3183 (2020).
28. N. Iwano, T. Adachi, K. Aoki, Y. Nakamura, M. Hamada, *Nat. Comput. Sci.* **2**, 378–386 (2022).
29. H. Altae-Tran, B. Ramsundar, A. S. Pappu, V. Pande, *ACS Cent. Sci.* **3**, 283–293 (2017).
30. J. Jiménez-Luna, F. Grisoni, G. Schneider, *Nat. Mach. Intell.* **2**, 573–584 (2020).
31. P. Das *et al*., *Nat. Biomed. Eng.* **5**, 613–623 (2021).
32. D. Nagarajan *et al*., *J. Biol. Chem.* **293**, 3492–3509 (2018).
33. W. Jin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2105070118 (2021).
34. F. Wong *et al*., *Nat. Commun.* **12**, 2321 (2021).
35. D. Lowe, "Why AlphaFold won't revolutionise drug discovery," *Chemistry World*, 5 August 2022; https://www.chemistryworld.com/opinion/why-alphafold-wont-revolutionise-drug-discovery/4016051.article.
36. F. Wong *et al*., *Mol. Syst. Biol.* **18**, e11081 (2022).
37. E. B. Breidenstein, C. de la Fuente-Núñez, R. E. Hancock, *Trends Microbiol.* **19**, 419–426 (2011).
38. A. H. Vo, T. R. Van Vleet, R. R. Gupta, M. J. Liguori, M. S. Rao, *Chem. Res. Toxicol.* **33**, 20–37 (2020).
39. N. Palmer, J. R. M. A. Maasch, M. D. T. Torres, C. de la Fuente-Nunez, *Infect. Immun.* **89**, e00703-20 (2021).
40. A. M. Watkins, R. Rangan, R. Das, *Structure* **28**, 963–976.e6 (2020).

41. N. E. Wheeler, P. P. Gardner, L. Barquist, *PLOS Genet.* **14**, e1007333 (2018).
42. H. Chen *et al.*, *Brief. Bioinform.* **22**, bbaa068 (2021).
43. D. Bojar, R. K. Powers, D. M. Camacho, J. J. Collins, *Cell Host Microbe* **29**, 132–144.e3 (2021).
44. D. Fisch *et al.*, *eLife* **8**, e40560 (2019).
45. B. Hie, E. D. Zhong, B. Berger, B. Bryson, *Science* **371**, 284–288 (2021).
46. P. J. Sample *et al.*, *Nat. Biotechnol.* **37**, 803–809 (2019).
47. E. Ong *et al.*, *Bioinformatics* **36**, 3185–3191 (2020).
48. G. W. Ashdown *et al.*, *Sci. Adv.* **6**, eaba9338 (2020).
49. L. Boeck *et al.*, *Nat. Microbiol.* **7**, 1431–1441 (2022).
50. N. Chaudhary, D. Weissman, K. A. Whitehead, *Nat. Rev. Drug Discov.* **20**, 817–838 (2021).
51. M. C. Crank *et al.*, *Science* **365**, 505–509 (2019).
52. M. Stracy *et al.*, *Science* **375**, 889–894 (2022).
53. D. M. Cornforth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5125–E5134 (2018).
54. J. Lee *et al.*, *Bioinformatics* **36**, 1234–1240 (2020).
55. H. C. Metsky *et al.*, *Nat. Biotechnol.* **40**, 1123–1131 (2022).
56. A. Khaledi *et al.*, *EMBO Mol. Med.* **12**, e10264 (2020).
57. E. S. Kavvas *et al.*, *Nat. Commun.* **9**, 4306 (2018).
58. C. Weis *et al.*, *Nat. Med.* **28**, 164–174 (2022).
59. X. Mei *et al.*, *Nat. Med.* **26**, 1224–1228 (2020).
60. M. Roberts *et al.*, *Nat. Mach. Intell.* **3**, 199–217 (2021).
61. D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, J. J. Collins, *Cell* **173**, 1581–1592 (2018).
62. L. F. de Lima, A. L. Ferreira, M. D. T. Torres, W. R. de Araujo, C. de la Fuente-Nunez, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106724118 (2021).
63. K. Pardee *et al.*, *Cell* **159**, 940–954 (2014).
64. K. Pardee *et al.*, *Cell* **165**, 1255–1266 (2016).
65. M. Karlikow *et al.*, *Nat. Biomed. Eng.* **6**, 246–256 (2022).
66. R. A. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25722–25731 (2020).
67. H. de Puig *et al.*, *Sci. Adv.* **7**, eabh2944 (2021).
68. N. M. Angenent-Mari, A. S. Garruss, L. R. Soenksen, G. Church, J. J. Collins, *Nat. Commun.* **11**, 5057 (2020).
69. J. A. Valeri *et al.*, *Nat. Commun.* **11**, 5058 (2020).
70. G. Chuai *et al.*, *Genome Biol.* **19**, 80 (2018).
71. H. K. Kim *et al.*, *Nat. Biotechnol.* **36**, 239–241 (2018).
72. D. Wang *et al.*, *Nat. Commun.* **10**, 4284 (2019).
73. E. Shrock *et al.*, *Science* **370**, abd4250 (2020).
74. V. Turbé *et al.*, *Nat. Med.* **27**, 1165–1170 (2021).
75. A. R. Wattam *et al.*, *Nucleic Acids Res.* **42**, D581–D591 (2014).

ACKNOWLEDGMENTS

We thank X. Tan, M. N. Anahtar, and J. A. Valeri for helpful comments on the manuscript. **Funding:** F.W. was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number K25AI168451. C.F.N. holds a Presidential Professorship at the University of Pennsylvania, is a recipient of the Langer Prize awarded by the AIChE Foundation, and acknowledges funding from the IADR Innovation in Oral Care Award, the Procter & Gamble Company, United Therapeutics, a BBRF Young Investigator Grant, the Nemirovsky Prize, a Penn Health-Tech Accelerator Award, the Dean's Innovation Fund from the Perelman School of Medicine at the University of Pennsylvania, the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM138201, and the Defense Threat Reduction Agency (HDTRA11810041, HDTRA1-21-1-0014, and HDTRA1-23-1-0001). J.J.C. was supported by the Defense Threat Reduction Agency (HDTRA12210032), the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R01-AI146194, and the Broad Institute of MIT and Harvard. This work is part of the Antibiotics-AI Project, which is directed by J.J.C. and supported by the Audacious Project; Flu Lab, LLC; the Sea Grape Foundation; and Rosamund Zander and Hansjorg Wyss for the Wyss Foundation. **Competing interests:** J.J.C. is scientific cofounder and scientific advisory board chair of EnBiotix, an antibiotic drug discovery company, and Phare Bio, a nonprofit venture focused on antibiotic drug development. C.F.N. provides consulting services to Invaio Sciences and is a member of the scientific advisory boards of Nowture S.L. and Phare Bio. F.W. declares no competing interests. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 1 April 2023; accepted 5 June 2023
10.1126/science.adh1114

REVIEW

The central role of density functional theory in the AI age

Bing Huang[1]*, Guido Falk von Rudorff[2,3]*, O. Anatole von Lilienfeld[4,5,6,7,8]*

Density functional theory (DFT) plays a pivotal role in chemical and materials science because of its relatively high predictive power, applicability, versatility, and computational efficiency. We review recent progress in machine learning (ML) model developments, which have relied heavily on DFT for synthetic data generation and for the design of model architectures. The general relevance of these developments is placed in a broader context for chemical and materials sciences. DFT-based ML models have reached high efficiency, accuracy, scalability, and transferability and pave the way to the routine use of successful experimental planning software within self-driving laboratories.

It is hardly controversial that we live in an era of artificial intelligence (AI). AI has touched upon and affected almost every branch of human activity, assuming center stage in many domains of daily life, such as natural language processing, computer vision, and forecasting. An AI-based robot scientist named "Adam" to assist and accelerate scientific discovery was introduced to the synthetic biology field nearly 20 years ago (*1*). Promising first steps with regard to robotic and autonomous experimentation for chemistry and materials, however (*2*, *3*), such as delivering self-driving laboratories for thin film discoveries (*4*), have only been made more recently. Nevertheless, such groundbreaking progress has already led Krenn *et al.* to survey community members and to fundamentally reconsider the meaning of "understanding" in the context of the scientific process itself (*5*). As outlined previously by Aspuru-Guzik *et al.* (*6*), the success of autonomous self-driving laboratories in chemistry and materials will depend crucially on the availability of machine learning (ML)–based control software capable of reliably forecasting and ranking experimental outcomes throughout the materials and chemical compound space (CCS) with sufficient accuracy in real time (*7*). CCS refers to the tremendously large set that emerges for all conceivable combinations of chemical composition, structures, and experimental conditions that result in stable forms of matter (*8*). Thermodynamic and kinetic stability being well defined through the quantum statistical mechanics of electrons and nuclei, reliance on a quantum mechanics–based approach toward CCS is an obvious way forward. Unfortunately, the relevant equations of quantum and statistical mechanics can only be solved exactly for the simplest of systems, rendering numerical solutions of approximate expressions necessary. Method development for the physics-based computational design and discovery of materials and molecules in CCS represents a long-standing challenge and has motivated decades of atomistic simulation research (*9*, *10*) Applications are as diverse as the atomistic sciences and include improved solutions for batteries, transistors, catalysts, coatings, ligands, alloys, and photovoltaics, among others. All such efforts have in common that they attempt to virtually navigate CCS to narrow down the search space for subsequent experimental verification and characterization.

The importance of electronic structure information for computational materials identification, characterization, and optimization has recently been highlighted by Marzari *et al.* (*11*). The most powerful compromise between predictive power and computational burden for calculating the properties and behavior of gaseous and condensed systems from first principles is density functional theory (DFT). In particular, the effective single-particle flavor of DFT, approximating the electronic kinetic energy contribution within the Kohn-Sham (KS) framework (*12*), has proven immensely useful. A countless number of ever-improving approximations to the exact exchange-correlation potential has led to further progress (*13*). With one of the cofounders of DFT, Walter Kohn, having been awarded the Nobel Prize in Chemistry in 1998, expectations for further improvements ran high at the turn of the century (*14*) and during the widespread adaption of ab initio molecular dynamics and time-dependent DFT to also treat thermal effects and electronic excitations, respectively. Consequently and unsurprisingly, two DFT-related contributions featured among the top 10 papers of all ti as highlighted in *Nature* in 2014 (*15*). An contemporary reviews have described further improvements (*16*), highlighted the importance of numerical reproducibility (*17*), or emphasized the importance of electron density as a measure of quality in addition to energy (*13*) (Fig. 1D). Among all of the relevant energy modes that affect the statistical mechanics of matter through their partition function (as manifested, for example, by translational, rotational, or vibrational degrees of freedom), calculating the electronic structure contribution represents the largest computational bottleneck, even when using DFT rather than more traditional post–Hartree Fock (post-HF) methods. In fact, the execution of DFT-based computational chemistry and materials protocols consumes major fractions of publicly available high-performance compute allocations (*18*) and has still largely failed to become a modern standard of industry 4.0 ("digital twins"). This situation is aggravated by CCS's steep combinatorial scaling and by neglect of correlations, i.e., relying on DFT-based compute campaigns that treat every system independently. Encouragingly, relying heavily on DFT, physics-based supervised quantum machine learning (QML) approaches have been introduced in recent years. This is not to be confused with the alternative use of QML when referring to the examination of quantum computing's potential to accelerate ML. Instead, we refer to QML specifically as classical ML of quantum observables regressed throughout chemical compound space. This use is well aligned with the convention of prefixing "quantum" within atomistic simulation, such as quantum Monte Carlo (QMC), quantum molecular dynamics, or even quantum chemistry. The encouraging, if not overwhelming, progress of physics-based QML in this domain originates, in our opinion, in its universal capability to generalize to such a degree that quantum observables can reliably be inferred for out-of-sample compounds across CCS as long as sufficient and representative data were provided for training.

It is difficult to overstate the general importance of these developments. In particular, as also argued previously (*19*), the emergence of generalizing statistical surrogate models (ML) indicates the formation of a fourth pillar in the hard sciences. This notion is universally applicable, going even beyond just the chemical and materials sciences. More specifically, the first, second, third, and fourth pillars correspond to manual experimentation, the theoretical framework to explain and predict experimental observables, numerical simulation tools for computationally complex equations of the theoretical framework, and statistical learning approaches that exploit relations encoded in experimental or simulated training data to infer observables, respectively. These pillars clearly build onto each other, and DFT can be seen as

[1]University of Vienna, Faculty of Physics, AT1090 Wien, Austria. [2]University Kassel, Department of Chemistry, 34132 Kassel, Germany. [3]Center for Interdisciplinary Nanostructure Science and Technology (CINSaT), 34132 Kassel, Germany. [4]Vector Institute for Artificial Intelligence, Toronto, Ontario M5S 1M1, Canada. [5]Department of Chemistry, University of Toronto, St. George Campus, Toronto, Ontario M5S 3H6, Canada. [6]Department of Materials Science and Engineering, University of Toronto, St. George Campus, Toronto, Ontario M5S 3E4, Canada. [7]Department of Physics, University of Toronto, St. George Campus, Toronto, Ontario M5S 1A7, Canada. [8]Machine Learning Group, Technische Universität Berlin and Berlin Institute for the Foundations of Learning and Data, 10587 Berlin, Germany.
*Corresponding author. Email: bing.huang@univie.ac.at (B.H.); vonrudorff@uni-kassel.de (G.F.v.R.); anatole.vonlilienfeld@utoronto.ca (O.A.v.L.)

bridging and encompassing them, all the way from the experimentally observable electron probability distribution through the Hohenberg-Kohn theorem (demonstrating the uniqueness of the electron density) and the Kohn-Sham ansatz (addressing the problem of the kinetic electronic energy) through the many numerical implementations and hardware use cases to training data generation and for informing the design of data-efficient ML model architectures. Figure 2 illustrates some of the possible ways that have already been explored to make use of DFT within ML models. To substantiate our view toward the key role that DFT plays in the fourth pillar of science, we will now review many of the specific ML contributions that have strongly benefited from it. To this end, this review is structured according to four categories, efficiency, accuracy, scalability, and transferability (EAST) (Fig. 2). EAST components represent an intuitive ordering principle that allows us to meaningfully discuss, distinguish, and compare some of the most important features necessary for building and using digital twins within the chemical and materials sciences.

Efficiency

Compared with DFT (or higher-level quantum chemistry), one of the most notable features of QML models is their unparalleled prediction speed after training. Although both approaches start from the same information entering the electronic Hamiltonian in the form of the external potential (i.e., chemical composition and geometry), ML model predictions are statistical surrogate model evaluations that amount to simple and efficient linear algebra operations, typically multiple orders of magnitude faster than DFT. By contrast, conventional physics-based simulators such as quantum calculations involve the computation of electronic integrals and iterative solvers of differential equations (typically diagonalization), both of which are high-dimensional, nonlinear, and computationally more demanding. However, to fully assess the efficiency of ML models, the associated cost for both, computational load for testing and training, and data acquisition has to be accounted for. Within the context of CCS, data are typically scarce, and data acquisition plays a crucial role. Training data needs are typically quantified in the form of learning curves (Fig. 1A), measuring the amount of reference data such as those coming from DFT calculations, which are required to reach a certain prediction error for out-of-sample queries, i.e., compounds that have not been part of training. Given sufficient data, the former point on model complexity becomes the numerical bottleneck. ML model complexity is roughly proportional to the number of parameters used for training and testing of models. Contrary to parametric neural network or random forest models, nonparametric ML models such as Gaussian process regression become less efficient as the number of training compounds increases. Furthermore, increasingly complex representations can reduce model efficiency. Considerable contemporary ML research is devoted to maximizing numerical efficiency through hardware (graphics processing units), as well as model architecture (optimizers and representations) (*20*).

Typically, there is a trade-off between a ML model's efficiency and its predictive power. For instance, ML models trained on fewer data can be more efficient but less accurate than those trained on more data. Sampling training data in representative ways still constitutes a substantial challenge when trying to reach universal accuracy already for small fractions of CCS (*21*). To better deal with this trade-off between efficiency and capability to generalize, one can rely on active learning schemes to sample training data more effectively than through random selection. Active learning (Fig. 3B) attempts to emulate the optimization of training set selection (*22*), for example, by biasing the training selection using query feature similarity (*23*, *24*). One can view this as an attempt to deal with the scaling of compound space that is closely related to the idea of "on-the-fly" learning (*25*). As a result, learning curves are steepened and data efficiency is improved.

Alternatively, one can make use of various types of Δ-ML (*26*) (refined in Fig. 2) to enhance predictive accuracy by learning corrections to labels, rather than absolute labels, thus exploiting accurately reflected trends in lower-level methods as well as error cancellation. For example, Δ-ML between DFT and coupled cluster singles, doubles, and perturbative triples [CCSD(T), a very accurate post-HF method] has been shown to yield corrected ab initio molecular dynamics trajectories with improved accuracy (*27*). Similarly, within transfer learning, optimized neural network weights of the first few layers trained on low-fidelity data (e.g., DFT) can be transferred to the model trained on high-fidelity data (e.g., CCSD) (*28*). Although effective in practice, transfer learning may suffer from low explainability.

Furthermore, going beyond mere Δ-ML (*26*, *29*) and transfer learning (*28*), correlations between multiple quantum approximations can be exploited through use of multilevel grid combination QML (CQML) approaches (*30*, *31*). CQML is analogous to composite quantum chemistry (often DFT based), or Jacob's ladder within DFT (Fig. 4), and allows for systematic error cancellations in hierarchical datasets of basis-set and electron correlation dimensions (Fig. 3B). For example, few high-fidelity data (e.g., double-hybrid DFT) for small systems, some medium-fidelity data (e.g., GGA) for medium-sized systems, and many low-fidelity data [e.g., local density approximation (LDA)] for large systems can be combined to yield high-fidelity quality predictions (Fig. 4). The CQML model in (*30*) combines three different dimensions, electron correlation, basis set, and training molecules. Within the realm of DFT, unifying amons (*24*) (i.e., small molecular fragments made up of typically no more than seven nonhydrogen atoms obtained through systematic fragmentation) and hierarchical density functionals, as demonstrated for the widely-known Jacob's ladder

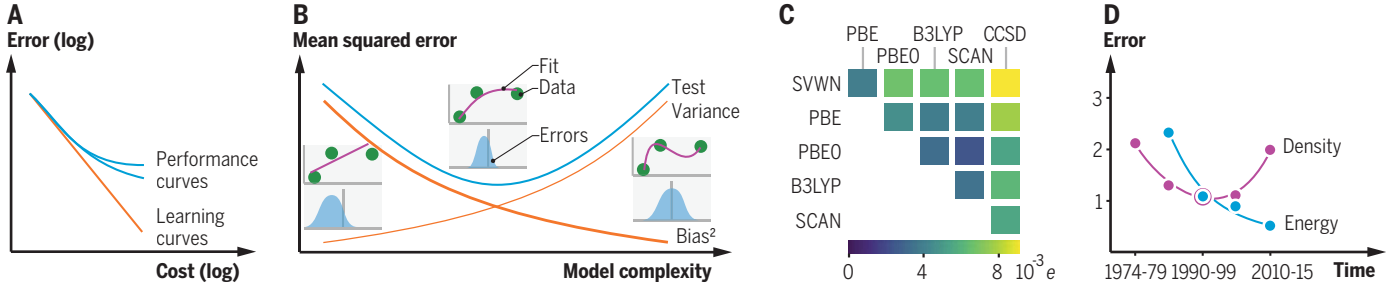


Fig. 1. Challenges for ML. (**A**) Prediction errors decay with inverse power laws in the limit of large training sets (*33*), whereas performance curves, i.e., related metrics not explicitly included in loss function, do not necessarily improve arbitrarily. (**B**) Test errors are a combination of a bias from the flexibility of the model to fit the data points and the variance arising from the model's flexibility between data points. (**C**) Integrated DFT-based electron density differences, a quantitative measure for density accuracy, of popular density functionals averaged over a random subset of QMrxn20 reactant molecules (*72*). (**D**) DFT errors of energy have improved over time yet yield lower accuracy for corresponding electron densities [average median-normalized error from literature (*13*)].

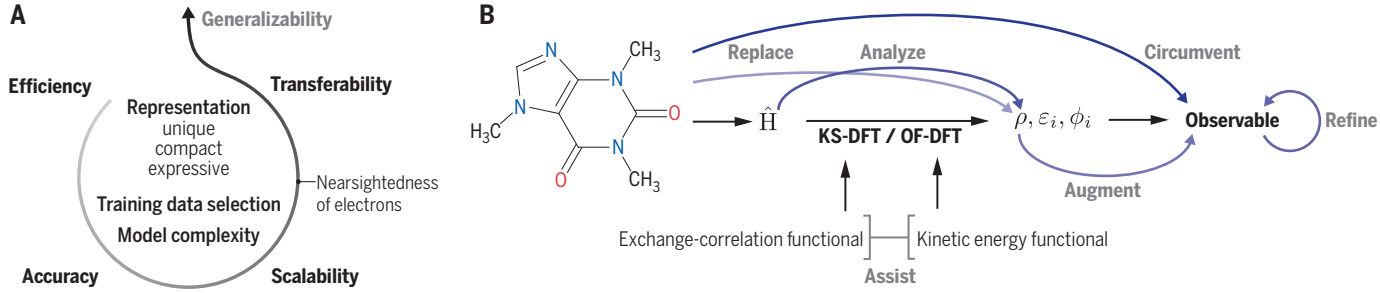


Fig. 2. Conceptual overview. (**A**) To review the key role of DFT, we rely on four key categories of predictive ML models of chemical and materials properties and processes, EAST: E (efficiency), A (accuracy), S (scalability), and T (transferability). A model of EAST generalizes well to unseen systems. (**B**) ML approaches going beyond DFT. Black indicates the traditional workflow in KS or orbital-free (OF) DFT in which a molecule defines the Hamiltonian, which yields the KS orbitals ϕ_i (or electron density, $\rho = \Sigma|\phi_i|^2$) and eigenvalues ε_i. Colors indicate ML-based routes taken to build upon DFT.

(Fig. 4), one can also greatly improve model efficiency with a set of extra, low-cost calculations [e.g., LDA]. Therefore, mixing DFT with other high-level electron correlation models such as CCSD(T) or QMC within the framework of CQML is a very promising strategy (*29*), particularly for the construction of large-scale datasets of potential energy surfaces. In addition to efforts to make DFT property-driven ML more efficient, another promising direction for improved cost reduction is to use ML to help improve orbital-free DFT, i.e., removing the explicit dependence on orbitals for the kinetic energy term in KS-DFT by ML the kinetic energy density functional (*32*) directly from data.

Highly efficient ML models enable us to minimize data needs and model complexity. Benefiting in such a way from adherence to Occam's razor, the resulting ML models will enable rapid iterations throughout CCS, virtually as well as in the real world through autonomous robotic execution and evaluation of experiments.

Accuracy

The underlying promise of all ML is that training on more reference data will result in more accurate models. Barring overfitting and inherent limitations due to extrapolation, any sufficiently flexible regressor converges the prediction error for smooth label functions down to noise level in the limit of infinite training data, which is consistent with the standard deviation of the error distribution decaying with the inverse square root of sample number for any fixed domain (central limit theorem). The leading term of prediction error typically decays according to an inverse power law for kernel methods as well as neural networks (*33*, *34*), and on log-log scales, learning curves (Fig. 1A) must therefore exhibit linearly decaying prediction errors as a function of increasing training set sizes. Consequently, learning curves play a crucial role for the comparative assessment of ML models. The systematic improvement of a model is only guaranteed for the learned label alone, whereas derived properties might improve coincidentally, yielding performance curves (Fig. 1A). This bears some analogy to DFT functionals, in which improvements in accuracy with respect to energies does not necessarily imply improvement of electron densities (*13*) (Fig. 1D). Note the importance of sufficient flexibility and converged cross-validation to achieve systematic results without overfitting (Fig. 3B). At constant training set size, ML models become more accurate when hard requirements and boundary conditions are accounted for, e.g., when including three- and four-body interactions (*35*) or by reducing delocalization errors (*36*). The representation (feature set) that either serves as the ML model input or is learned by the ML model itself is crucial. As mentioned and reviewed on multiple occasions (*7*, *37*), training data–efficient representations should be unique or complete (in the sense of a bijective, one-to-one relationship to the external potential), compact (small size of feature vector), and sufficiently expressive to account for underlying physics such as power-law relations, symmetries, invariances, or constraints. A major outstanding challenge is quantifying uncertainty (*38*), especially because residual errors for common tasks are distinctly non-normal (*39*). Similar ideas have been successful (e.g., Jacob's ladder; Fig. 4) within DFT with LDA, PBE, or SCAN (strongly constrained and appropriately normed) satisfying known constraints for DFT functionals (*40*), which has helped the accuracy of DFT despite its use of uncontrolled approximations, with only few bounds on the discretization error available (*41*).

Most DFT-based ML models introduced so far roughly fall into any one of the following three categories (Fig. 2).

Learning of electronic observables

Learning of electronic observables, i.e., quantum mechanical averages for which the electronic details of density or wavefunctions have been integrated out (circumvented/replaced in Fig. 2), relies on input based on the same information as for the model Hamiltonian: atomic coordinates, nuclear charges, number of electrons, and multiplicity. Once trained, direct QML (predicting atomic forces, molecular energies, or other quantum properties) is analogous to black-box usage of DFT and is the most common method used so far. Because chemical compound space formally grows combinatorically with the $4N$ dimensions manifested by chemical elements and atomic positions, the main drawback of direct learning is that it will generally fall short of the desired accuracy because of the exorbitant training data needs. For example, despite years of active research, the QM9 challenge to reach chemical accuracy when estimating atomization energies for simple organic molecules in gas phase (*42*) after training on just 100 examples has not yet been solved.

Recently introduced hybrid ML/DFT approaches

Recently introduced hybrid ML/DFT approaches improve the DFT model construction by learning the effective Hamiltonian as an intermediate quantity (replaced in Fig. 2), from which target properties follow straightforwardly, e.g., SchNorb (*43*) or DeepH (*44*). In addition to improving accuracy, this hybrid approach can offer better accuracy for intensive properties such as the energy gap between the HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital). Similarly, ML strategies have improved semiempirical quantum chemistry (*45*) and tight-binding DFT (*46*, *47*).

ML density functionals

ML density functionals (Assist in Fig. 2) come in two variants. The first is within the KS-DFT framework and improves mapping from the electron density to the exchange-correlation energy based on higher-level reference data [e.g., DFT/CCSD density and CCSD(T) energy]. With this method, it is possible to approximate the exact unknown exchange-correlation functional with high accuracy. The main drawback, however, is that the computational cost associated with the explicit dependence on orbitals in the kinetic energy and exchange term is not removed. Well-known contributions relying on this method include NeuralXC (*48*) and DM21 (*36*). The second strategy is more in the spirit

of Hohenberg-Kohn, building the orbital-free map from electron density to energy (augmented in Fig. 2). First shown for the kinetic energy density functional (*49*), this line of research also includes the map from potential to density to energy (*50*, *51*).

Numerical approximations in DFT are rooted in careful neglect of certain physical effects. To reach experimental accuracy, ML models might eventually even require the inclusion of experimental observables to guarantee improvement over synthetic computational approaches, e.g., by automated and data-driven approaches (*52*). Given the strategies outlined, we believe that ML on DFT is reaching a point where reliable forecasting of materials' properties is possible, but autonomous experimental exploration will be necessary to further improve accuracy and applicability.

Scalability

Scalability is critical for enabling the study of larger and more complex electronic systems. Although linear scaling DFT-based implementations such as ONETEP have made great progress (*53*), DFT generally scales cubically with system size, which is considerably more favorable than accurate post-HF methods such as CCSD(T) [~$O(N^7)$]. Nevertheless, the routine use of ab initio molecular dynamics simulations using the most accurate flavors of DFT, hybrid or range-separated DFT in particular, becomes rapidly elusive for larger systems, for example, in small proteins such as ubiquitin (Fig. 4). ML can also help in this regard, most commonly by partitioning models of extensive properties onto atomic contributions.

The scalability of ML models can also be assessed in a more chemical sense, namely its ability to generalize to larger query systems after training on smaller systems only. This rests upon the locality assumption that is implied when using similarity measurements based on atomic environments. This assumption is often justified with reference to Kohn's nearsightedness of electronic matter (*54*, *55*). For example, as pointed out by Kohn and others (*56*, *57*), the locality of the one-particle electron density matrix of a system with periodic boundary condition decays exponentially with the HOMO/LUMO gap. Nearsightedness has a series of consequences, such as the extent of charge transfer (which has bigger impact on charged species), conjugation, electron correlation (in particular for strongly correlated systems), and London dispersion (in large biomolecules as well as noncovalently interacting molecular complexes) (Fig. 2).

In practice, a scalable ML model has the necessary but not sufficient requirement that it coarse-grains the representation such that negligible long-range interactions can be omitted. Improving scalability ranks perhaps among the most common and severe challenges of state-of-the-art ML models, possibly because of the difficulty of accurately including long-range interactions. Among the many long-range effects, interacting Coulombic multipole moments and their polarizabilities have been considered first within multiscale ML models (*58*) because of their well-known classical structure and the availability of many empirical models. As for other long-range effects of quantum mechanical origin, such as conjugation, spin-orbit coupling, surface crossing, and electron correlation in general, robust scalable ML-based models are yet to be devised.

Once the short- and long-range effects accessible through DFT are properly accounted for by scalable ML models, the routine study of condensed systems, macromolecules, defects, and maybe even grain boundaries will become affordable to the entire community.

Transferability

The physical origins of chemical transferability of atoms, bonds, or functional groups have been much studied and are closely related to the aforementioned locality and nearsightedness of electronic matter (*55*). Here, we associate transferability predominantly with the capability to generalize short-range effects across CCS. Long-range effects are often less subtle and complex by comparison and matter more in the context of scalability, as discussed before. Examples for short-range effects include well-known assignment of interatomic many-body contributions in terms of force-field topologies such as the degrees of freedom associated with covalent bonding, angles, or torsion, and steric hindrance. As illustrated by the amon concept (*24*) (Fig. 4) for various labels calculated with DFT, these projections of internal degrees of freedom, together with their off-equilibrium distortions for every conformer, can be conveniently represented throughout chemical space by molecular fragments with systematically increasing size. Other effects, dominated by long-range characteristics, can be left for approximate treatments with empirical or other relevant models (*59*). As a recent example of the insight gained from such a transferable approach, as implemented by training on both local small bottom-up and large top-down fragments, we refer to the stark differences observed between smooth and accurate DFT-trained ML trajectories of the protein crambin in aqueous solution and classical force field–based Brownian motion–like counterparts with stochastic characteristics (*60*).

ML practitioners not familiar with DFT might not be aware of the fact that different quantum properties exhibit strongly varying degrees of transferability depending on their exact definition. The transferability of energy, for instance, depends on whether it is the total energy (derived from a particular theoretical level), an orbital's eigenvalue, an atom's energy within a molecule, the electron correlation energy, reorganization or relaxation energies, or the relative energy between two theoretical levels. For example, the latter point was shown recently, indicating that ML models of molecular HF energies require more training data than energies calculated with methods that include electron correlation such as DFT or QMC, possibly because, on average,

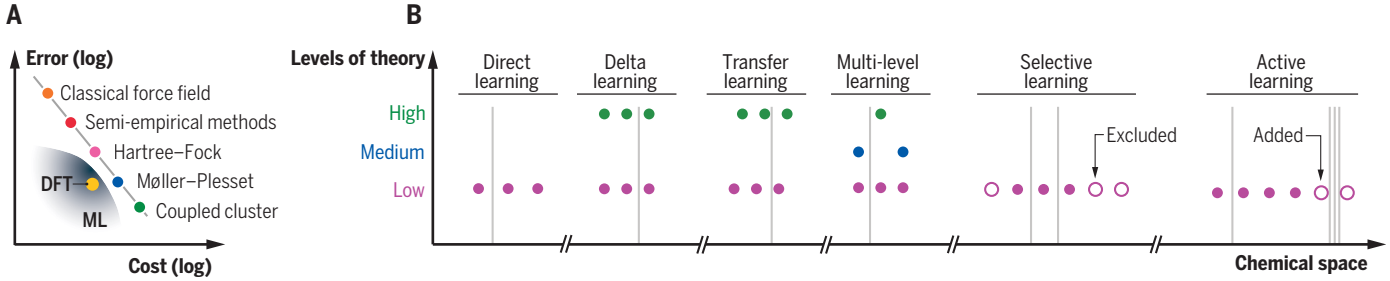


Fig. 3. Budget-aware computing strategies for sampling chemical compound space. (**A**) Reference data acquisition cost accuracy Pareto front with popular atomistic simulation approaches. At mean-field cost, DFT stands out for approaching the accuracy of explicit electron correlation methods. (**B**) Overview of selection strategies for increased data efficiency when training models are capable of sampling chemical compound space. Some include training data (points) from different levels of theory (colored) resulting in models to be queried for new compounds (gray lines). Direct learning refers to all data points from one level, delta and transfer learning approximate differences between levels on aligned and disjoint training systems, respectively. Multilevel learning exploits hierarchical data structures across multiple levels. Selective learning trains on data points (filled) nearest to queries, ignoring all others (empty). Active learning biases the data collection (empty) to improve overall prediction accuracy.

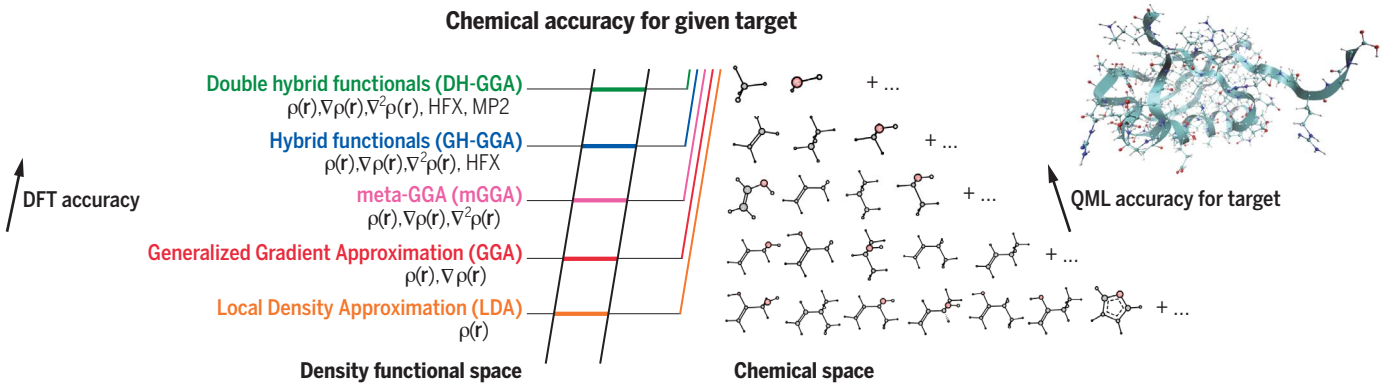


Fig. 4. Jacob's ladder (left) refers to the hierarchy of increasingly less approximate parameterizations commonly made to model the exact unknown exchange-correlation potential. Used within CQML and in combination with hierarchies in fragment space, e.g., amons (right) (*24*), properties of large query targets (e.g., the small protein ubiquitin shown as an inset) can be predicted. For each pair of amon subset (ranked by system size) and density functional approximation (ranked by Jacob's ladder), the computational burden for training data generation is correlated. First demonstrated for organic chemistry amons and GGAs, hybrid functionals, and QMC in (*29*) HFX, HF exchange; MP2, Møller Plesset second order.

correlation tends to smoothen the energy function by contraction of atomic radii (*29*). Conversely, consider the increase in transferability (and consequently reduction in data needs) for absolute energies versus energy differences between two levels of theory as exemplified for DFT and a higher-level post-HF method for Δ-ML that was demonstrated in 2015 (*26*).

Although early studies mostly focused on proof-of-principle and atomic and interatomic energetics throughout CCS, more recent research indicates that properties at the finer electronic resolution can be transferable as well. For instance, transferable ML models of atomic nuclear magnetic resonance shifts calculated with DFT were published in 2015 (*61*). Further, the electron density of small molecules such as ethene or butadiene was found to be transferable to octatetraene (*62*). Learning the deformation density (*63*) can also help to improve the transferability of electron density. Further evidence of transferability at electronic resolution includes the interelectronic force field approach (*64*) and the localized molecular orbital feature–based approach (*65*). Superior model-transfer ability can also arise from the use of electronic features such as Mulliken charges and bond order (*66*).

Another important and fundamental concept at the electronic level, the density functional (DF), has been deemed transferable. Approximated DFs constructed from (biased) heuristics, however, are prone to severe transferability issues. Such shortcomings, also described by Nagai *et al.* (*67*), could be alleviated by an in-depth analysis of prediction error distributions (*39*), which would lead to a more systematic generation of DFs using data-driven ML approaches (assisted in Fig. 2) (*49*). By incorporating additional physical constraints into ML, one can enhance the transferability of DFs. This has also been shown more recently by learning the nonlocal exchange DF with a tailored representation that preserves the density distribution under uniform scaling (*68*). Beyond DFT, physical constraints (such as Kato's cusp conditions) are also crucial in the correlated framework, because they help to enhance the expressiveness of the deep neural network for wavefunction approximation (*69*).

High transferability across CCS represents the ultimate test for DFT as well as ML. Encouraging progress has been made so far, indicating the possibility of sampling CCS more freely and concurrently paving the way toward software control solutions that will routinely be able to handle even exotic chemistries and formulations, e.g., within the self-driving laboratory setting.

Conclusions

We have reviewed the instrumental role that DFT has played in the emergence of QML-based models that enable the navigation of chemical compound space with EAST. In addition to its fundamental theoretical underpinnings in terms of a quantum mechanical approximation method, DFT has served as a truly outstanding source of calculated properties for freely chosen molecules or materials with controllable and reasonable acquisition costs and most welcome accuracy. Outstanding electronic structure challenges to DFT-based QML include surface crossing, open-shell and spin-orbit coupling effects, conductivity, and excited states dynamics. However, DFT can already be extremely useful as an intermediate-quality method that can be exploited in multifidelity QML models or for inspiration of physics-based QML architectures.

The availability of large, diverse, and high-accuracy (at the experimental level or above) materials and molecular property datasets will remain a fundamental requirement for the development and training of transferable QML models that can universally handle any property and chemistry and that can conveniently be incorporated within the experimental planning software of future self-driving and closed-loop autonomous experimentation. We consider the general scarcity and lack of data to represent one of the most severe current roadblocks for our community's path toward that goal. For example, many available databases report only select properties and molecular graphs for Lewis structure–conforming systems. Notable exceptions, such as DFT-based distortions along normal modes and conformers reported in QM7-X (*70*), are few and scattered throughout the chemical compound space. Representative data encoding transition states, defects, charged species, radicals, entire ab initio molecular dynamics trajectories, *d*- and *f*-elements, or excited states, to name just some, are still mostly lacking.

Although notable achievements have been made in successfully applying physics-based ML to DFT solutions throughout chemical and materials sciences, there remains a dearth of theoretical research focused on the underlying fundamentals. Some of the basic open questions include: (i) Can one rigorously define CCS in a mathematical way, akin to the Hilbert space for electronic wavefunctions, to quantify its inherent properties such as dimensionality, density, and volume? (ii) CCS is discrete in reality, yet which maps enable smooth interpolations into latent spaces that facilitate inverse design (*71*)? (iii) Can there be a single QML model that allows for a unified yet accurate description of any chemical compound, regardless of its size, composition, aggregation state, and external conditions? In summary and roughly speaking, DFT has affected QML as an ab initio solver, with ever-improving performance of ML-based exchange-correlation and/or kinetic energy density functionals, as a hybrid DFT/ML framework for building effective Hamiltonians, and as a robust computational workhorse for generating highly relevant and affordable synthetic data. Reflecting on the four EAST categories,

efficiency, accuracy, scalability, and transferability, it is clear that DFT has played a pivotal role in bridging all of the pillars of modern science, from experiments through theory and simulation to physics-based ML model building. Seeing the impressive progress made by building onto DFT, we do not think that it is far-fetched to expect these developments to directly lead to the development and widespread adaptation of autonomous experimentation. Therefore, building on the four preceding pillars, we might very well witness the emergence of the next, fifth pillar of science in the not-too-distant future: self-driving laboratories throughout the hard sciences! Apart from the above-described progress, noteworthy key components of such self-driving laboratories will include the seamless integration of physics-based QML models with robotic hardware and devices to perform the various experimental tasks in a laboratory, such as the preparation of reactants and ingredients, the execution of the experiment, or subsequent purification and analysis of the resulting products. Although they are not yet deployable for arbitrary chemistries and experiments, because of their numerical efficiency and reliability, DFT-based QML models will likely play a crucial role in achieving this integration.

REFERENCES AND NOTES

1. R. D. King *et al*., *Nature* **427**, 247–252 (2004).
2. S. V. Ley, D. E. Fitzpatrick, R. J. Ingham, R. M. Myers, *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).
3. J. M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, *Nature* **559**, 377–381 (2018).
4. B. P. MacLeod *et al*., *Sci. Adv.* **6**, eaaz8867 (2020).
5. M. Krenn *et al*., *Nat. Rev. Phys.* **4**, 761–769 (2022).
6. A. Aspuru-Guzik, R. Lindh, M. Reiher, *ACS Cent. Sci.* **4**, 144–152 (2018).
7. O. A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, *Nat. Rev. Chem.* **4**, 347–358 (2020).
8. P. Kirkpatrick, C. Ellis, *Nature* **432**, 823 (2004).
9. G. Ceder, *Science* **280**, 1099–1100 (1998).
10. N. Marzari, *Nat. Mater.* **15**, 381–382 (2016).
11. N. Marzari, A. Ferretti, C. Wolverton, *Nat. Mater.* **20**, 736–749 (2021).
12. W. Kohn, L. J. Sham, *Phys. Rev.* **140** (4A), A1133–A1138 (1965).
13. M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew, K. A. Lyssenko, *Science* **355**, 49–52 (2017).
14. A. E. Mattsson, *Science* **298**, 759–760 (2002).
15. R. Van Noorden, B. Maher, R. Nuzzo, *Nature* **514**, 550–553 (2014).
16. K. Burke, *J. Chem. Phys.* **136**, 150901 (2012).
17. K. Lejaeghere *et al*., *Science* **351**, aad3000 (2016).
18. S. Heinen, M. Schwilk, G. F. von Rudorff, O. A. von Lilienfeld, *Mach. Learn. Sci. Technol.* **1**, 025002 (2020).
19. O. A. von Lilienfeld, *Angew. Chem. Int. Ed.* **57**, 4164–4169 (2018).
20. N. J. Browning, F. A. Faber, O. Anatole von Lilienfeld, *J. Chem. Phys.* **157**, 214801 (2022).
21. P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, A. Michaelides, *J. Chem. Phys.* **153**, 034702 (2020).
22. N. J. Browning, R. Ramakrishnan, O. A. von Lilienfeld, U. Roethlisberger, *J. Phys. Chem. Lett.* **8**, 1351–1359 (2017).
23. J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A. E. Roitberg, *J. Chem. Phys.* **148**, 241733 (2018).
24. B. Huang, O. A. von Lilienfeld, *Nat. Chem.* **12**, 945–951 (2020).
25. G. Csányi, T. Albaret, M. C. Payne, A. De Vita, *Phys. Rev. Lett.* **93**, 175503 (2004).
26. R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
27. M. Bogojeski, L. Vogt-Maranto, M. E. Tuckerman, K.-R. Müller, K. Burke, *Nat. Commun.* **11**, 5223 (2020).
28. J. S. Smith *et al*., *Nat. Commun.* **10**, 2903 (2019).
29. B. Huang, O. A. von Lilienfeld, J. T. Krogel, A. Benali, *J. Chem. Theory Comput.* **19**, 1711–1721 (2023).
30. P. Zaspel, B. Huang, H. Harbrecht, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **15**, 1546–1559 (2018).
31. R. Batra, G. Pilania, B. P. Uberuaga, R. Ramprasad, *ACS Appl. Mater. Interfaces* **11**, 24906–24918 (2019).
32. K. Ryczko, S. J. Wetzel, R. G. Melko, I. Tamblyn, *J. Chem. Theory Comput.* **18**, 1122–1128 (2022).
33. C. Cortes, L. D. Jackel, S. Solla, V. Vapnik, J. Denker, "Learning curves: Asymptotic values and rate of convergence," in *Advances in Neural Information Processing Systems 6* (NIPS, 1993); pp. 327–334.
34. K. R. Müller, M. Finke, N. Murata, K. Schulten, S. Amari, *Neural Comput.* **8**, 1085–1106 (1996).
35. B. Huang, O. A. von Lilienfeld, *J. Chem. Phys.* **145**, 161102 (2016).
36. J. Kirkpatrick *et al*., *Science* **374**, 1385–1389 (2021).
37. B. Huang, O. A. von Lilienfeld, *Chem. Rev.* **121**, 10001–10036 (2021).
38. M. Reiher, *Isr. J. Chem.* **62**, (2021).
39. P. Pernot, B. Huang, A. Savin, *Mach. Learn. Sci. Technol.* **1**, 035011 (2020).
40. J. Sun, A. Ruzsinszky, J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
41. E. Cancès, G. Dusson, G. Kemlin, A. Levitt, *SIAM J. Sci. Comput.* **44**, B1312–B1340 (2022).
42. O. A. von Lilienfeld, K. Burke, *Nat. Commun.* **11**, 4895 (2020).
43. K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R. J. Maurer, *Nat. Commun.* **10**, 5024 (2019).
44. H. Li *et al*., *Nat. Comput. Sci.* **2**, 367–377 (2022).
45. P. O. Dral, O. A. von Lilienfeld, W. Thiel, *J. Chem. Theory Comput.* **11**, 2120–2125 (2015).
46. J. J. Kranz, M. Kubillus, R. Ramakrishnan, O. A. von Lilienfeld, M. Elstner, *J. Chem. Theory Comput.* **14**, 2341–2352 (2018).
47. M. Stöhr, L. Medrano Sandonas, A. Tkatchenko, *J. Phys. Chem. Lett.* **11**, 6835–6843 (2020).
48. S. Dick, M. Fernandez-Serra, *Nat. Commun.* **11**, 3509 (2020).
49. J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, *Phys. Rev. Lett.* **108**, 253002 (2012).
50. Y. Kiat, Y. Vortman, N. Sapir, *Nat. Commun.* **10**, 2540 (2019).
51. Y. Bai, L. Vogt-Maranto, M. E. Tuckerman, W. J. Glover, *Nat. Commun.* **13**, 7044 (2022).
52. S. Steiner *et al*., *Science* **363**, eaav2211 (2019).
53. J. C. A. Prentice *et al*., *J. Chem. Phys.* **152**, 174111 (2020).
54. E. Prodan, W. Kohn, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11635–11638 (2005).
55. S. Fias, F. Heidar-Zadeh, P. Geerlings, P. W. Ayers, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11633–11638 (2017).
56. R. Baer, M. Head-Gordon, *Phys. Rev. Lett.* **79**, 3962–3965 (1997).
57. W. Kohn, *Phys. Rev. Lett.* **76**, 3168–3171 (1996).
58. A. Grisafi, J. Nigam, M. Ceriotti, *Chem. Sci.* **12**, 2078–2090 (2020).
59. T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Nat. Commun.* **12**, 398 (2021).
60. O. T. Unke *et al*., Accurate machine learned quantum-mechanical force fields for biomolecular simulations. arXiv: 2205.08306 [physics.chem-ph] (2022).
61. M. Rupp, R. Ramakrishnan, O. A. von Lilienfeld, *J. Phys. Chem. Lett.* **6**, 3309–3313 (2015).
62. A. Grisafi *et al*., *ACS Cent. Sci.* **5**, 57–64 (2019).
63. K. Low, M. L. Coote, E. I. Izgorodina, *J. Chem. Theory Comput.* **18**, 1607–1618 (2022).
64. M. Cools-Ceuppens, J. Dambre, T. Verstraelen, *J. Chem. Theory Comput.* **18**, 1672–1691 (2022).
65. T. Husch, J. Sun, L. Cheng, S. J. R. Lee, T. F. Miller 3rd, *J. Chem. Phys.* **154**, 064108 (2021).
66. C. Duan, A. Nandy, H. Adamji, Y. Roman-Leshkov, H. J. Kulik, *J. Chem. Theory Comput.* **18**, 4282–4292 (2022).
67. R. Nagai, R. Akashi, O. Sugino, *NPJ Comput. Mater.* **6**, 43 (2020).
68. K. Bystrom, B. Kozinsky, *J. Chem. Theory Comput.* **18**, 2180–2192 (2022).
69. J. Hermann, Z. Schätzle, F. Noé, *Nat. Chem.* **12**, 891–897 (2020).
70. J. Hoja *et al*., *Sci. Data* **8**, 43 (2021).
71. R. Gómez-Bombarelli *et al*., *ACS Cent. Sci.* **4**, 268–276 (2018).
72. G. F. von Rudorff, S. N. Heinen, M. Bragato, O. A. von Lilienfeld, *Mach. Learn. Sci. Technol.* **1**, 045026 (2020).

ACKNOWLEDGMENTS

Funding: O.A.v.L. has received support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 772834) and as the Ed Clark Chair of Advanced Materials and Canada CIFAR AI Chair. **Competing interests:** The authors declare no competing interests.

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

Submitted 27 April 2023; accepted 30 May 2023
10.1126/science.abn3445

EDITORIAL

New options for patenting in Europe

Inventors seeking to protect their intellectual property in Europe now have an important choice to make. More than a decade after its approval, a new patent system came into effect last month, which some European Union (EU) countries already have signed on to. Inventors in these countries can use either the new unitary patent (UP) system or the previously existing route of European patents. A balance between the two paths is hoped to spur a more robust pan-European patent landscape. However, the new system must be widely adopted first.

Historically, patent protection was available for European jurisdictions through national patent systems. Since 1978, the European Patent Office (EPO) has offered a harmonized application and examination path in the now 39 signatory states to the European Patent Convention. But as patents granted by the EPO ultimately still become national patent rights that are maintained and litigated nationally, the system is fragmented and costly relative to other regions.

In 2012, the EU voted in favor of a process that contains a European patent with unitary effect, a language regime, and a single patent jurisdiction, the Unified Patent Court. In June, the new patent system was initiated in 17 states, including Germany, France, and Italy, although the UK, Poland, and Switzerland have chosen not (yet) to join. By allowing inventors to apply for a single patent that is recognized across all UP states, the new system (operated by the EPO) will reduce patent owners' costs.

In addition, the Unified Patent Court intends to harmonize the legal treatment of patent rights in UP countries. It will have exclusive jurisdiction over both unitary patents and European patents. Although all active European patents were to be automatically transferred into the new system, more than half a million (539,665) such patents have, for the moment, opted out, as permitted during a transitional period (of up to 12 years).

Which patent path should inventors choose? A unitary patent will be attractive to those seeking protection in a large number of countries of the European Patent Convention. Patent seekers operating in one or two national markets are unlikely to want a unitary patent. For sectors where international patent portfolios are large, such as pharmaceuticals, chemicals, and electronics, the new system is a likely choice. Smaller and/or financially constrained actors, such as universities, small and medium-sized enterprises, and start-ups, may opt for selective strategies, seeking unitary patent protection only for particularly promising inventions.

For those who pursue a European patent, yet another choice must be made—whether to opt out of the Unified Patent Court. Opting out would make patents subject to national infringement and invalidation decisions and would avoid having unfavorable decisions apply across all member countries of the new system. If the number of current patent holders who opt out is an early indicator of trust, then it appears that the new scheme is being viewed cautiously at best.

> "...the new system... raises the specter of an even more fragmented European patent process."

Still, the overall impact of the new system is hard to assess at this point. In general, the lower costs of unitary patents are likely to invite more filings. However, the low patent renewal fees (especially during the first 10 years) could make it more attractive for lower-quality inventions to seek patent protection. This puts more burden on the EPO to weed out such applications. The new system may also favor so-called patent assertion entities who could use the threat of pan-European injunctions to extract more money from genuine inventors. Strategic behavior, such as forum shopping, may become more pronounced as well. For example, the Unified Patent Court, with its large number of court locations, allows plaintiffs to choose courts, which may produce favorable first-instance rulings. Given the high costs of litigation in the new system, these choices can influence negotiation positions and outcomes, possibly to the detriment of small enterprises, start-ups, and research organizations.

The EPO and the Unified Patent Court will ultimately determine the balance between the strength of patent rights and patent quality in the new system. Although the new system seeks harmonization across countries, it unfortunately raises the specter of an even more fragmented European patent process. Although large patent portfolio holders will benefit the most from lower fees, smaller entities may be more hampered by the increasing complexity. Overall, there is a fair chance that the new, long-awaited unitary patent system may miss its target of providing better protection for high-quality inventions.

–Reinhilde Veugelers and Dietmar Harhoff

Reinhilde Veugelers is a professor at KU Leuven, Leuven, Belgium; a senior fellow at Bruegel, Brussels, Belgium; and a senior fellow at the Peterson Institute for International Economics, Washington, DC, USA. reinhilde.veugelers@kuleuven.be

Dietmar Harhoff is a director at the Max Planck Institute for Innovation and Competition, Munich, Germany. dietmar.harhoff@ip.mpg.de

10.1126/science.adj6400

NEWS

> “You’re not going to improve by 27%. You may worsen a little slower than what you might have expected.”
>
> **Harvard University neurologist Alvaro Pascual-Leone,** in Endpoints News, about the limitations of lecanemab, a drug fully approved last week by U.S. regulators for early symptoms of Alzheimer’s disease.

IN BRIEF

Edited by **Jeffrey Brainard**

A draft law in Europe would deregulate gene-edited crops such as disease-resistant potatoes.

AGRICULTURE

Europe considers loosening rules on gene-edited crops

The European Commission last week proposed making it easier to study and commercialize gene-edited plants, engineered to have beneficial traits such as resistance to climate change, diseases, and pests. Scientists welcomed the draft bill, but it could take the European Parliament and Council years to approve the measure. It would exempt gene-edited crops from lengthy risk assessments required for genetically modified crops, with traits created by transferring DNA from other species, under several conditions. For example, gene-edited crops must be equivalent to what conventional plant breeding could produce and have no more than 20 nucleotide base pairs added or replaced. Researchers may also add or move the same genes as conventional breeders. Like the current rules, the draft law prohibits gene-edited plants in organic agriculture and requires gene-edited seeds to be labeled. It also prohibits gene editing that adds herbicide resistance, reflecting calls in Europe to reduce use of herbicides. If approved, the European Union’s policy would largely match those of China, the United Kingdom, and the United States.

U.N. sets emissions goals for ships

CLIMATE POLICY | The United Nations’s International Maritime Organization (IMO) last week set a nonbinding target for international cargo ships to emit no net greenhouse gas emissions by about 2050. The target is more ambitious than IMO’s first climate goal, set in 2018, which called for halving emissions by 2050. If the shipping industry achieves the control targets, it could help limit global warming to no more than 2°C, according to an analysis by the International Council on Clean Transportation, a think tank. But environmental groups complain that the shipping goal lacks enforcement mechanisms. They also point out that the warming target exceeds 1.5°C, a threshold beyond which scientists foresee severe impacts. IMO, which regulates international shipping through its member states, will encourage the design of more energy-efficient cargo ships and the use of fuels with lower carbon emissions by 2030. IMO will re-evaluate the strategy in 5 years.

New Chinese journal gains impact

PUBLISHING | Just 3 years after launch, *The Innovation*, a China-based, English language journal, has made a splash. Last month, it notched a citation impact factor of 32.1—behind only *Nature*'s 64.8 and *Science*'s 56.9 among multidisciplinary journals—in the annual Journal Citation Reports released by the Clarivate analytics company. A group of young Chinese scientists pooled their savings to get the open-access journal up and running, according to the *South China Morning Post*, which first reported the story. Despite producing nearly 30% of the world's reviewed scientific papers, China has produced few highly ranked journals. At least two-thirds of *The Innovation*'s published papers come from China-based corresponding authors. It accepts only about 13% of submissions.

Judge pauses Mexico's science law

POLICY | Researchers who have complained that a new law in Mexico could overly centralize the nation's funding and policies for science won this week a preliminary court victory, the full impact of which was unclear as *Science* went to press. A district judge issued an order that provisionally suspends the law, enacted in May, pending a final court decision. Hundreds of scientists had filed legal motions asserting that the law violates their rights including to intellectual property and freedom of research. The new national science agency created by the law, the National Council for the Humanities, Sciences, Technologies and Innovation, says it has not received the court decision but will appeal any ruling that prevents implementation of the law.

Cooling speeds orchid evolution

EVOLUTION | Periods of global cooling spurred the rapid speciation of ground-dwelling orchids, a study has found, suggesting global changes in climate have systematically contributed to the generation of biodiversity. Researchers examined 1475 species of Orchidoideae, a subfamily that contains more than 3500 species. They found that new species did not emerge gradually or steadily; instead, the speciation rate was much faster during global cooling events. Other factors such as latitude and elevation did not significantly affect the rate. Reporting this week in the *Proceedings of the National Academy of Sciences*, the researchers say these orchids, found on every continent besides Antarctica, provide one of the clearest examples of long-term changes in climate influencing species diversification. The authors say it's not clear why cooling aided speciation in this group of plants, but interbreeding between populations may have been disrupted during colder periods.

Owl monkeys were research subjects at a Colombian lab before it was closed and lost U.S. funding.

PHOTO: PEOPLE FOR THE ETHICAL TREATMENT OF ANIMALS

ANIMAL RESEARCH

NIH ends funding for troubled monkey lab

The U.S. National Institutes of Health (NIH) has canceled funding eligibility for a malaria research facility in Colombia following allegations that research animals were abused. The decision in late June about the Caucaseco Scientific Research Center came 3 months after an investigation by the animal rights group People for the Ethical Treatment of Animals spurred Colombian authorities to close the facility and seize more than 100 monkey test subjects, including ones reportedly confined to rusted and feces-encrusted cages. The shutdown coincided with a report from the U.S. Government Accountability Office that criticized NIH for lax oversight of animal care at foreign facilities, including a lack of site visits. The husband-and-wife team that runs the facility has received $17.6 million in funding from NIH since 2003.

AI yields designer proteins

STRUCTURAL BIOLOGY | The artificial intelligence (AI) program DALL-E has dazzled the public by instantly converting text such as "teddy bears working underwater" into digital images. Now, a new AI program offers similar help for researchers seeking novel proteins tailored for specific tasks. The new program, called RFdiffusion, allows researchers to upload a 3D model of a cell surface receptor associated with a disease, for example. The software can then identify the amino acid sequences and shapes of proteins likely to bind to the receptor, providing candidates for drug testing. Compared with previous protein design programs not powered by AI, RFdiffusion increases accuracy as much as 100-fold, a team of U.S. researchers reports this week in *Nature*. That could help scientists create proteins useful for a range of tasks, such as harvesting atoms of precious metals from seawater and extracting carbon dioxide from the atmosphere.

Satellites leak radio waves

ASTRONOMY | Radiation leaking from SpaceX's internet-beaming Starlink satellites is polluting parts of the radio spectrum, potentially interfering with ground-based radio telescopes, a study has found. Using the Low Frequency Array (LOFAR) in the Netherlands, researchers detected "unintended electromagnetic radiation" from the onboard electronics of 47 out of 68 observed Starlink satellites, unrelated to internet signals or other downlinks, the research team reported last week in *Astronomy & Astrophysics*. Radio astronomers had previously raised the alarm that Starlink downlinks could interfere with protected radio bands used to identify organic molecules in space. With more than 2000 Starlinks in orbit and many more planned, the authors urged satellite operators and regulators to consider these emissions in future designs.

IN DEPTH

Muds in Crawford Lake capture humanity's fossil fuel burning, fertilizer use, and atomic bomb fallout.

GEOLOGY

Anthropocene's emblem may be Canadian pond

Proposal to mark global changes of the 1950s now faces skeptical scientific bureaucracy

By **Paul Voosen**

The Anthropocene—the moment in geological time when humanity became a world-altering force—is a major step closer to formal recognition. A group of stratigraphers announced this week it will nominate muds from Crawford Lake, a 24-meter-deep pond in Ontario in Canada, to serve as the "golden spike" marking the beginning of a distinct Anthropocene epoch. Chemicals and minerals in the sediments clearly capture the global "great acceleration" in fossil fuel burning, fertilizer use, and atomic bomb fallout that began in the 1950s.

The proposal must now win votes in the coming months from three layers of the geological bureaucracy, and approval is not a given. The transformation the Crawford muds capture is not in question; similar changes are detectable in lakes, oceans, ice, peat bogs, and corals worldwide. "It's not rocket science to see that the planet has been operating under different rules than before," says Francine McCarthy, a micropaleontologist at Brock University who led work on the proposed site. But geologists are deeply divided about whether to recognize the Anthropocene as a distinct epoch.

For 14 years, the Anthropocene Working Group, established by the International Commission on Stratigraphy (ICS), the body that governs the geologic timescale, has studied whether and when the world has departed the Holocene, the period of stable climate that began 11,700 years ago at the end of the last ice age, and entered a new epoch defined by humanity's global influence. In 2016, the group voted that such a departure occurred in the 1950s with postwar economic growth and a surge in fossil fuel use (*Science*, 26 August 2016, p. 852).

But the proposal still needed an exemplar, a single geologic deposit that best captures the changes, just as a layer of iridium between two clays in a Moroccan cliffside left by the dinosaur-killing asteroid strike marks the end of the Cretaceous period 66 million years ago. For the past half-decade, 12 sites have vied to be the Anthropocene's golden spike (*Science*, 6 May 2022, p. 562).

Crawford Lake sediments feature annual bands of white calcite that correspond to human activity.

The choice came down to Crawford and Sihailongwan Lake, in a protected forest in northeastern China. Both lakes are deep, so sediments have settled to the bottom and accrued in undisturbed layers for centuries. But Crawford's sediment bands are thicker and feature annual bands of white calcite crystals, which precipitate out of the lake during summer algae blooms fed by nutrients from fertilizer runoff and other human activities. "Without people peeing and pooping and luring geese, living there and impacting the land, we don't see that reliable annual signature," McCarthy says. The muds are also rich in fly ash pollutants and contain trace amounts of plutonium released by atmospheric nuclear weapons tests, among other signatures.

In the next few months, the working group will submit three proposals for a vote by the Subcommission on Quaternary Stratigraphy (SQS), the ICS division that oversees the recent past. The first names Crawford Lake as the golden spike, and the second offers supporting evidence from several other sites, perhaps including Sihailongwan. The third and most contentious proposes formal recognition of the Anthropocene as an epoch. The process is highly political, says Hernán Bobadilla, a geologist and philosopher of science at the Polytechnic University of Milan. "This is not science beyond doubt."

While the geologists have deliberated about how to define the Anthropocene, other scientists, as well as artists, writers, and politicians, have already seized on the term. Some use it for the start of any human influence on Earth, beginning with the dawn of agriculture 12,000 years ago. Others use the term to mark changes that began several thousand years ago, as humans developed the ability to clearcut forests and build huge structures. "Everyone wants the Anthropocene for their own purposes," says Martin Head, a Brock geologist and vice chair of SQS.

Formalizing the Anthropocene as an epoch will clarify matters, says Colin Waters, a geologist at the University of Leicester and chair of the working group. But Philip Gibbard, a geologist at the University of Cambridge and secretary-general of ICS, worries that by pinning the Anthropocene to the 1950s, ICS will turn off many who have taken the term to refer to different periods.

Gibbard and others have instead proposed that the Anthropocene be considered an informal "event," which does not require an exact global starting point in time. "It's more inclusive and more representative of what has happened over a long time, as people have spread around the Earth," says Dorothy Merritts, a geomorphologist at Franklin & Marshall College.

Tensions are running high. Waters says the working group has been told not to talk in advance with SQS members, and the group's announcement of Crawford Lake was not allowed to happen at the International Congress on Stratigraphy, taking place this week in Lille, France, but rather at a nearby hotel. Influential members who will vote on the proposal have also taken prominent stances for and against it: SQS is led by two proponents, Head and Jan Zalasiewicz, a Leicester paleobiologist who first chaired the working group, whereas Gibbard, a detractor, is influential at ICS, which would vote after SQS.

If the proposals pass ICS, they would be submitted for final approval from the International Union of Geological Sciences, which is headed by a longtime opponent, Stanley Finney, a stratigrapher at California State University, Long Beach.

It's possible that the proposal for a separate Anthropocene epoch will fail, but the lake's record of human impact could mark a new "Crawfordian" stage within the existing Holocene epoch, Waters says. "We just don't know how it will be viewed." If the submission falters, it can't be proposed again for a decade. By that time, some of the resistant stratigraphers may move on—but the signals of the Anthropocene will remain, Head says. "We won't be going back to the Holocene envelope. It won't be just a blip." ■

WORKFORCE

Abortion bans and more disrupt retention and recruitment

Amid U.S. political landscape, some academics face tough choices about where to work

By **Katie Langin**

After the U.S. Supreme Court overturned the constitutional right to an abortion in June 2022, Alesha Thayer sat down with her husband and worked out a plan to leave the state where she has been a faculty member for more than 5 years. Abortions had immediately become outlawed there, and Thayer, who has experienced miscarriages and complicated pregnancies and hopes to grow her family, feared the new legal landscape would limit her reproductive autonomy and the care she could receive. "We just can't be here anymore," the physical scientist says. (Thayer is not her real name; *Science* granted anonymity to multiple sources for this story so they could speak candidly.)

> **"There's so many great places and, academically and intellectually, I hate the prospect of being limited."**
>
> **Tara Jenson**, University of Wisconsin-Milwaukee

Thayer is not alone. "I have a lot of friends who are in the same boat … trying to get out," she says. Department chairs in states that have restricted abortion or adopted other conservative initiatives targeting transgender rights or diversity, equity, and inclusion (DEI) efforts have started to receive resignation notices, a *Science* investigation has found. "We lost one faculty member who was ready to leave Texas, and two others we retained acknowledged that they were exploring other schools due to frustration with Texas politics," said a chair in the state, which allows its citizens to sue anyone who assists an abortion and passed a law last month banning university DEI initiatives.

Such laws do not seem to have made a sizable dent in the number of applicants for faculty or graduate student positions, according to *Science*'s survey of 40 STEM departments in states with restrictive abortion laws. Of the 21 that responded, only four reported a recent drop. Many more reported increases, especially from prospective international students—a change that likely reflects the easing of the pandemic.

Multiple departments found, however, that some prospective students declined offers because of state laws that made them feel unwelcome. Even in departments that haven't yet experienced impacts, chairs worry about the future. "I can't imagine that this will not have an impact on us recruiting talent at all levels," one said.

Science did not include postdoc positions in its survey because hiring is typically done by individual labs, not the department, which makes data much harder to collect. But some faculty members report trouble there as well. "I have had women that are interested in potentially doing a postdoc with me tell me that they will not come to Alabama—I've had that said multiple times," says Amy Weinmann, an immunology professor at the University of Alabama at Birmingham (UAB) who published an opinion piece in September 2022 discussing how the new abortion laws may impact women's health care options. She understands why postdocs would be wary of moving somewhere like Alabama, where abortion is illegal except when a pregnant person's health is seriously at risk. For one, abortion drugs can be a lifesaving treatment for someone experiencing pregnancy complications. "You can't just get on a plane and deal with that," she notes.

In addition to the potential impacts on recruiting early-career scientists, Weinmann worries about leadership positions. When UAB's School of Medicine searched for a department chair and a dean position last year, a total of nine men and zero women were given on-campus interviews—a sign that few women may have applied, she says. "These demographics would be unheard of at a major academic medical center a year ago, yet we saw them in two consecutive leadership searches." (UAB Provost and Senior Vice President of Academic Affairs Pam Benoit says "every leadership search we conduct involves a diverse search committee and seeks interest from any and all qualified candidates.")

While the geologists have deliberated about how to define the Anthropocene, other scientists, as well as artists, writers, and politicians, have already seized on the term. Some use it for the start of any human influence on Earth, beginning with the dawn of agriculture 12,000 years ago. Others use the term to mark changes that began several thousand years ago, as humans developed the ability to clearcut forests and build huge structures. "Everyone wants the Anthropocene for their own purposes," says Martin Head, a Brock geologist and vice chair of SQS.

Formalizing the Anthropocene as an epoch will clarify matters, says Colin Waters, a geologist at the University of Leicester and chair of the working group. But Philip Gibbard, a geologist at the University of Cambridge and secretary-general of ICS, worries that by pinning the Anthropocene to the 1950s, ICS will turn off many who have taken the term to refer to different periods.

Gibbard and others have instead proposed that the Anthropocene be considered an informal "event," which does not require an exact global starting point in time. "It's more inclusive and more representative of what has happened over a long time, as people have spread around the Earth," says Dorothy Merritts, a geomorphologist at Franklin & Marshall College.

Tensions are running high. Waters says the working group has been told not to talk in advance with SQS members, and the group's announcement of Crawford Lake was not allowed to happen at the International Congress on Stratigraphy, taking place this week in Lille, France, but rather at a nearby hotel. Influential members who will vote on the proposal have also taken prominent stances for and against it: SQS is led by two proponents, Head and Jan Zalasiewicz, a Leicester paleobiologist who first chaired the working group, whereas Gibbard, a detractor, is influential at ICS, which would vote after SQS.

If the proposals pass ICS, they would be submitted for final approval from the International Union of Geological Sciences, which is headed by a longtime opponent, Stanley Finney, a stratigrapher at California State University, Long Beach.

It's possible that the proposal for a separate Anthropocene epoch will fail, but the lake's record of human impact could mark a new "Crawfordian" stage within the existing Holocene epoch, Waters says. "We just don't know how it will be viewed." If the submission falters, it can't be proposed again for a decade. By that time, some of the resistant stratigraphers may move on—but the signals of the Anthropocene will remain, Head says. "We won't be going back to the Holocene envelope. It won't be just a blip." ■

WORKFORCE

Abortion bans and more disrupt retention and recruitment

Amid U.S. political landscape, some academics face tough choices about where to work

By **Katie Langin**

After the U.S. Supreme Court overturned the constitutional right to an abortion in June 2022, Alesha Thayer sat down with her husband and worked out a plan to leave the state where she has been a faculty member for more than 5 years. Abortions had immediately become outlawed there, and Thayer, who has experienced miscarriages and complicated pregnancies and hopes to grow her family, feared the new legal landscape would limit her reproductive autonomy and the care she could receive. "We just can't be here anymore," the physical scientist says. (Thayer is not her real name; *Science* granted anonymity to multiple sources for this story so they could speak candidly.)

> **"There's so many great places and, academically and intellectually, I hate the prospect of being limited."**
>
> **Tara Jenson**, University of Wisconsin-Milwaukee

Thayer is not alone. "I have a lot of friends who are in the same boat … trying to get out," she says. Department chairs in states that have restricted abortion or adopted other conservative initiatives targeting transgender rights or diversity, equity, and inclusion (DEI) efforts have started to receive resignation notices, a *Science* investigation has found. "We lost one faculty member who was ready to leave Texas, and two others we retained acknowledged that they were exploring other schools due to frustration with Texas politics," said a chair in the state, which allows its citizens to sue anyone who assists an abortion and passed a law last month banning university DEI initiatives.

Such laws do not seem to have made a sizable dent in the number of applicants for faculty or graduate student positions, according to *Science*'s survey of 40 STEM departments in states with restrictive abortion laws. Of the 21 that responded, only four reported a recent drop. Many more reported increases, especially from prospective international students—a change that likely reflects the easing of the pandemic.

Multiple departments found, however, that some prospective students declined offers because of state laws that made them feel unwelcome. Even in departments that haven't yet experienced impacts, chairs worry about the future. "I can't imagine that this will not have an impact on us recruiting talent at all levels," one said.

Science did not include postdoc positions in its survey because hiring is typically done by individual labs, not the department, which makes data much harder to collect. But some faculty members report trouble there as well. "I have had women that are interested in potentially doing a postdoc with me tell me that they will not come to Alabama—I've had that said multiple times," says Amy Weinmann, an immunology professor at the University of Alabama at Birmingham (UAB) who published an opinion piece in September 2022 discussing how the new abortion laws may impact women's health care options. She understands why postdocs would be wary of moving somewhere like Alabama, where abortion is illegal except when a pregnant person's health is seriously at risk. For one, abortion drugs can be a lifesaving treatment for someone experiencing pregnancy complications. "You can't just get on a plane and deal with that," she notes.

In addition to the potential impacts on recruiting early-career scientists, Weinmann worries about leadership positions. When UAB's School of Medicine searched for a department chair and a dean position last year, a total of nine men and zero women were given on-campus interviews—a sign that few women may have applied, she says. "These demographics would be unheard of at a major academic medical center a year ago, yet we saw them in two consecutive leadership searches." (UAB Provost and Senior Vice President of Academic Affairs Pam Benoit says "every leadership search we conduct involves a diverse search committee and seeks interest from any and all qualified candidates.")

For those seeking posts, the calculations can be challenging. Faculty positions in particular are highly competitive, so limiting one's options by ruling out certain states may be a significant career risk. At the Ph.D. and postdoc level, too, research opportunities can be difficult to come by, and junior researchers have no control over where their dream lab may be.

But at least some are choosing to prioritize state law in their searches. "There's so many great places and, academically and intellectually, I hate the prospect of being limited," says Tara Jenson, an environmental health epidemiologist who recently completed her Ph.D. at the University of Wisconsin-Milwaukee. But she and her wife don't want to live in a place that restricts the rights of women and those in the LGBTQ+ community, so she won't even consider postdoc or faculty jobs in some states. "They could offer me … an amazing deal and I still would be like, 'Nope, I will settle for less at a place that I feel safe and protected.'"

Mae Guthman, a neuroscience postdoc at Princeton University who is applying for faculty positions, feels similarly. She is interested in developing novel forms of gender-affirming hormone therapy—work that could be affected by some recently enacted state laws that restrict what medical care options are available for transgender people. "It's definitely impacting the institutions I apply to." Guthman—who is transgender—is also weighing her own health care needs and those of people she'd like to recruit to her lab. "Ideally, I would like to have … trans trainees," she says, "as well as birthing-capable trainees and trainees of color." So, she wants to be in a place where all such potential lab members feel welcome and safe.

Many department chairs in states with restrictive laws feel that their hands are tied. "We can control the atmosphere within the department and the values that the department expresses," one chair said. But they can't change what health care options are available in their state or influence how the state is perceived by the outside world. Speaking about Texas's DEI bill, one chair noted it "sends a message [that] we're not an inclusive environment."

At public universities, administrators face additional constraints as state employees, which limits what they are allowed to say. "To have a university that hasn't really talked about this, from my perspective, is dehumanizing," says a faculty member who works at a public university in a state that bans abortions. Last year—after more than 10 years at her current institution—she made the difficult decision to apply for faculty jobs elsewhere. "I actually have a written offer from another university," the biologist says. "My department chair does not know."

As for Thayer, she's still on the job market. "I'm waiting for the right fit to come along," she says. "It's hard at my stage. I have a very expensive lab. I have established my career already. It will be terribly disruptive to move." But she's determined to make it happen—so much so that if she doesn't land another faculty job within the next few years, she's prepared to leave academia. ■

In states where abortion is restricted, some professors have resigned and others are weighing their options.

FUNDING

Giant NSF grants aim to revitalize left-behind regions

34 teams make first cut for $160 million awards to drive economic and social change

By **Jeffrey Mervis**

Tara Hudiburg worries she may be "out of her league" in vying for the most lucrative competitive grant the National Science Foundation (NSF) has ever awarded. But the forest ecologist at the University of Idaho may be underestimating herself. Last month, her team's proposal to build a "smart climate economy" in the five-state Columbia River Basin was chosen as one of 34 semifinalists for NSF's first five Regional Innovation Engine (RIE) grants to be awarded this fall.

The $160 million RIE grants are the flagship initiative of NSF's new technology directorate. They are designed to broaden the agency's traditional focus on fundamental research to include turning discoveries into new industries and training workers for the resulting high-tech jobs. The grants are also supposed to improve economic conditions in what NSF calls "those regions of America that have not fully participated in the technology boom of the past several decades."

Many lawmakers believe NSF hasn't done enough to address those disparities, pointing out that five states receive almost 40% of its research dollars. So Congress will be watching closely to see whether the agency does a better job of spreading the wealth with this first cohort of winners. Hoping to reassure lawmakers, NSF opted for maximum transparency by identifying the semifinalists, a sharp break from tradition for proposals that are still under review.

Some applicants fear it isn't going far enough. For 4 decades NSF has run the Established Program to Stimulate Competitive Research (EPSCoR), which is designed to help smaller and more rural states boost their research capacity. But institutions in EPSCoR states such as Idaho are leads for only eight—fewer than one-quarter—of the RIE semifinalists, whose proposals span

For those seeking posts, the calculations can be challenging. Faculty positions in particular are highly competitive, so limiting one's options by ruling out certain states may be a significant career risk. At the Ph.D. and postdoc level, too, research opportunities can be difficult to come by, and junior researchers have no control over where their dream lab may be.

But at least some are choosing to prioritize state law in their searches. "There's so many great places and, academically and intellectually, I hate the prospect of being limited," says Tara Jenson, an environmental health epidemiologist who recently completed her Ph.D. at the University of Wisconsin-Milwaukee. But she and her wife don't want to live in a place that restricts the rights of women and those in the LGBTQ+ community, so she won't even consider postdoc or faculty jobs in some states. "They could offer me … an amazing deal and I still would be like, 'Nope, I will settle for less at a place that I feel safe and protected.'"

Mae Guthman, a neuroscience postdoc at Princeton University who is applying for faculty positions, feels similarly. She is interested in developing novel forms of gender-affirming hormone therapy—work that could be affected by some recently enacted state laws that restrict what medical care options are available for transgender people. "It's definitely impacting the institutions I apply to." Guthman—who is transgender—is also weighing her own health care needs and those of people she'd like to recruit to her lab. "Ideally, I would like to have … trans trainees," she says, "as well as birthing-capable trainees and trainees of color." So, she wants to be in a place where all such potential lab members feel welcome and safe.

Many department chairs in states with restrictive laws feel that their hands are tied. "We can control the atmosphere within the department and the values that the department expresses," one chair said. But they can't change what health care options are available in their state or influence how the state is perceived by the outside world. Speaking about Texas's DEI bill, one chair noted it "sends a message [that] we're not an inclusive environment."

At public universities, administrators face additional constraints as state employees, which limits what they are allowed to say. "To have a university that hasn't really talked about this, from my perspective, is dehumanizing," says a faculty member who works at a public university in a state that bans abortions. Last year—after more than 10 years at her current institution—she made the difficult decision to apply for faculty jobs elsewhere. "I actually have a written offer from another university," the biologist says. "My department chair does not know."

As for Thayer, she's still on the job market. "I'm waiting for the right fit to come along," she says. "It's hard at my stage. I have a very expensive lab. I have established my career already. It will be terribly disruptive to move." But she's determined to make it happen—so much so that if she doesn't land another faculty job within the next few years, she's prepared to leave academia. ■

In states where abortion is restricted, some professors have resigned and others are weighing their options.

FUNDING

Giant NSF grants aim to revitalize left-behind regions

34 teams make first cut for $160 million awards to drive economic and social change

By **Jeffrey Mervis**

Tara Hudiburg worries she may be "out of her league" in vying for the most lucrative competitive grant the National Science Foundation (NSF) has ever awarded. But the forest ecologist at the University of Idaho may be underestimating herself. Last month, her team's proposal to build a "smart climate economy" in the five-state Columbia River Basin was chosen as one of 34 semifinalists for NSF's first five Regional Innovation Engine (RIE) grants to be awarded this fall.

The $160 million RIE grants are the flagship initiative of NSF's new technology directorate. They are designed to broaden the agency's traditional focus on fundamental research to include turning discoveries into new industries and training workers for the resulting high-tech jobs. The grants are also supposed to improve economic conditions in what NSF calls "those regions of America that have not fully participated in the technology boom of the past several decades."

Many lawmakers believe NSF hasn't done enough to address those disparities, pointing out that five states receive almost 40% of its research dollars. So Congress will be watching closely to see whether the agency does a better job of spreading the wealth with this first cohort of winners. Hoping to reassure lawmakers, NSF opted for maximum transparency by identifying the semifinalists, a sharp break from tradition for proposals that are still under review.

Some applicants fear it isn't going far enough. For 4 decades NSF has run the Established Program to Stimulate Competitive Research (EPSCoR), which is designed to help smaller and more rural states boost their research capacity. But institutions in EPSCoR states such as Idaho are leads for only eight—fewer than one-quarter—of the RIE semifinalists, whose proposals span

A team from academia and industry hopes to win a grant to lower carbon emissions from Louisiana's oil and chemical industries.

PHOTO: DREW ANGERER/GETTY IMAGES

14 topical areas ranging from agriculture to sustainable energy to wellness. That means Hudiburg's team, which includes two tribal nations and a community college, must also compete against 26 proposals from 14 non-EPSCoR states, many led by top 20 research universities and major academic medical centers with large companies as partners.

One is headed by the world-renowned Cleveland Clinic and includes two major research universities, hospitals, and Medtronic, a leader in medical devices and therapeutics. That team proposes to build and operate a comprehensive digital platform to improve health care outcomes in the Cleveland and Detroit areas. Although neither Ohio nor Michigan is an EPSCoR state, neurologist Lara Jehi, the clinic's chief research information officer and co-leader on the proposal, notes that "Ohio and Michigan are among the worst nationally for health equity."

But some contenders believe having such juggernauts vie for the prize isn't consistent with the goals of the program. "The idea is not to award more money to wealthy institutions for work that they're already doing," says Michaele Armstrong, associate director of a regional hub for biotech startups at Washington State University that is part of the Idaho proposal. "Instead, an NSF Engine is supposed to help mobilize technologies or communities that have been left behind."

Some EPSCoR states contending for an NSF Engine are also home to well-resourced research institutions. A semifinalist team led by the University of Alabama at Birmingham (UAB), whose research portfolio ranks in the top 5% nationally, wants to use biotechnology to drive innovation and reduce health inequities plaguing the region's sizable Black population. UAB hopes to do that by partnering with some of the 14 historically Black colleges and universities (HBCUs) in the state.

"Alabama has the largest number of HBCUs in the country, and we also have a strong track record of recruiting minority patients in clinical trials," says Kathy Nugent, a biotech entrepreneur who oversees commercialization efforts at UAB and is leading the team. "But reducing health inequities is a hard problem, and the validation and visibility that comes with winning an NSF Engine would really help."

Another EPSCoR-state semifinalist, based at Louisiana State University (LSU), aims to address another urgent societal problem: the massive carbon footprint of the state's oil, gas, and chemical industry. "If we can figure out a cheap energy source, whether that's solar, nuclear, or wind, and then find a way to use its energy to transport and sequester our greenhouse gas emissions, we can essentially still be home to a very hydrocarbon intense industry," says Andy Maas, director of LSU's office of innovation and commercialization, who leads the team.

NSF expects the winning teams to also attract significant funding from both the public and private sectors. LSU should have no trouble satisfying that goal, Maas says: The big oil companies have announced some $27 billion in new low-carbon projects in the region, he notes, and New Orleans recently won $50 million in federal infrastructure funding to pursue green hydrogen.

At the core of every proposal is a plan to carry out and commercialize the research conducted by the team, then train a local workforce needed for the relevant business sector to thrive. Typically, commercialization involves helping a faculty member start a company or license their intellectual property. But a semifinalist based at the University of Texas at El Paso (UTEP)—in a state not eligible for EPSCoR—is proposing a new model that will allow small and medium-size companies in the aerospace and defense sector to order up the advances they need to prosper.

"Most of those companies cannot afford the innovation that comes from a major university or a big corporate lab," says Ahsan Choudhuri, a UTEP aerospace engineer who leads the team. "So we're offering innovation as a service. We'll provide them with the infrastructure, talent, and innovation to rebuild America's defense industrial base." He calls it "a very grassroots approach."

Every RIE proposal must also include plans to increase diversity, equity, and inclusion (DEI) in the U.S. research enterprise. That could be a challenge in states, such as Texas, which have ordered universities to roll back or shutter such efforts. But like the other semifinalists from such states who spoke with *Science*, Choudhuri does not anticipate any problem in satisfying that requirement.

"DEI is in our DNA," Choudhuri says, citing the ethnic and gender diversity of his leadership team. "Hispanics make up 80% of our community, so we don't have to set up an office to ensure diversity. By supporting us, you are helping to solve the problem."

NSF chose the first round of semifinalists from 188 proposals and will winnow the list further after site visits. Each of the five winners will receive $15 million for 2 years, with the promise of $145 million more over a decade. (The number and size of subsequent cohorts will depend on the agency's future overall budget.)

Hudiburg hopes to be one of the winners. But she won't think of herself as a loser if her team's proposal comes up short.

"The connections we've made are permanent," she says. "We have a dream, and if we don't get an NSF engine, we'll keep looking for other funding. Because the atmosphere can't wait." ■

A new funding agency could boost science progra at Indian universities, such as the University of Delhi.

RESEARCH FUNDING

India plans to invigorate science with hefty new funding agency

Proposed National Research Foundation would spend $6 billion over 5 years, but is drawing mixed reactions

By **Pratik Pawar**

The Indian government has announced an ambitious plan to create a new National Research Foundation (NRF) that would pump $6 billion into research over 5 years. But it is drawing mixed reviews from scientists.

Although few details are available, many researchers are welcoming the proposal to create a new "apex body" for Indian research, saying it could help boost the nation's relatively meager investments in basic and applied science. But some worry the foundation might fall prey to political interference. They are also skeptical it will be able to attract the envisioned funding, 70% of which is supposed to come from private industry. That goal is likely "a little unrealistic," says policy specialist Shailja Vaidya Gupta, a former senior science adviser to the government.

The NRF proposal, unveiled on 28 June by India's Union Council of Ministers, is rooted in a 2019 report requested by Prime Minister Narendra Modi. The report, from the Prime Minister's Science, Technology and Innovation Advisory Council, noted that India lagged far behind other large nations in the share of gross domestic product (GDP) it spends on research and in the quality of research papers and patents it produces. Many Indian universities, it added, do not conduct any research.

To help reverse those trends, the report called for a powerful new research agency, similar to the National Science Foundation in the United States, which would help coordinate science policy and consolidate and expand funding, including peer-reviewed grants to academic researchers. The agency's annual budget should equal 0.1% of India's GDP, or more than $2 billion at the time, it said. To protect the agency from political pressure and bureaucratic red tape, the report recommended that an independent board of prominent scientists choose its leaders, and that it operate with "complete autonomy … so that it does not face hurdles in funding good projects."

The new proposal departs substantially from that vision. It calls for the prime minister to preside over NRF's board, for example, and for the government's science and education ministers to fill two other top jobs. The science ministry and India's science adviser will also help oversee NRF and its governing board.

Having the prime minister occupy such a senior role "tells you the seriousness [with] which we want to take the R&D ecosystem," says Ajay Sood, the government's principal science adviser. And researchers say the structure means the foundation could prosper under politicians who support its mission. "It's not a bad idea for the prime minister to be at the helm," says Partha Majumder, founder of the National Institute of Biomedical Genomics and a member of India's Science and Engineering Research Board (SERB), created in 2008 in an earlier effort to address funding struggles.

NRF's links to powerful politicians could also help it attract industry support, Majumder says. Senior government officials "can probably talk to industry much more effectively than a bunch of us," he says.

But the arrangement could expose NRF to political meddling, warns the Breakthrough Science Society, an advocacy group. It notes that Modi's government has sometimes promoted fringe, pseudoscientific ideas—such as claims that ancient Hindus invented aircraft and the internet. That "cannot but make scientists apprehensive of the future direction and fate of science research under an NRF with such political leaders at its helm," the society said in a statement.

India's notoriously ponderous bureaucracy could pose another threat to the new agency, says Gupta, who worked on the NRF proposal in 2020 when she was a government official. Because the foundation would not be an independent agency, it will "have to adhere to rules which apply across sectors," she says. "So, if you're building a bridge, or you're discovering new proteins, you are treated the same way."

The plan calls for NRF to subsume SERB. Ultimately, the government says the foundation will receive more public money than it gave to SERB. But, overall, officials are counting on the private sector to contribute more than $4 billion of NRF's funding from 2023 through 2028.

Sood acknowledges that "these are aspirational numbers." But he says they reflect the government's desire "to bring other people [on board and] make them co-partners in the R&D and innovation ecosystem."

Still, others are skeptical. "Banking on the corporate sector to pitch in with a sum exceeding [the] government commitment is a wild dream," the Breakthrough Science Society said.

It's been difficult to fully assess the NRF proposal, researchers say, because the government will not publicly release the legislation creating the foundation until it is passed by India's parliament. "It shouldn't be this way … [the draft bill] should be up for public debate before we could put it up into the parliament," Gupta says.

Action on the bill could come as early as this summer, with Parliament scheduled to hold its Monsoon Session from 20 July to 11 August. ■

Pratik Pawar is a journalist in Bangalore, India.

SCIENTIFIC COMMUNITY

Scientists mourn Arecibo's impending shutdown

Astronomers scramble to find work as storied observatory faces uncertain future

By **Claudia López Lloreda**, *in Arecibo, Puerto Rico*

At the end of a winding road lined by flowering red flamboyan trees, you can still see the broken remains of the world's largest single-dish radio telescope. The Arecibo Observatory's 305-meter telescope collapsed in 2020, and in October 2022 its owner, the National Science Foundation (NSF), announced that after 6 decades, the site would no longer be an observatory. Instead, NSF plans to transform it into an educational facility. Contracts for up to 90 staff members, including 25 researchers, end on 14 August.

"We're closing out all science activities," says physicist Julie Brisset of the University of Central Florida (UCF), who's in charge of maintenance and research at Arecibo. To scientists who have worked at the observatory, used its data, or mentored local students, it is a painful transition. "It's very sad," says Brazilian physicist Pedrina Terra dos Santos, who has worked at the observatory for 17 years and, like many of her colleagues, is still seeking a new job.

Arecibo was once central to astronomy; just a few weeks ago, astronomers relied on archived data from its giant radio telescope in their report on the long-sought hum of gravitational waves from supermassive black holes. Even after engineering failures and poor maintenance brought down the big scope, Arecibo continued to observe the skies with its remaining instruments, including a 12-meter radio telescope and an optical lab for probing atmospheric particles and winds.

In October 2022, NSF announced it would not rebuild the giant telescope, saying it was following community recommendations for the best use of scarce research dollars. It is now shutting down most of the smaller instruments as well. As scientists depart, "all the expertise associated with instruments is leaving," Brisset says. Olga Figueroa-Miranda, director of the observatory, says people from UCF, Puerto Rico's Metropolitan University, and Yang Enterprises, an engineering firm, will be let go, including herself. She has yet to find a new position.

Angel Vazquez, an engineer and director of telescope operations at Arecibo, found a new job with the National Radio Astronomy Observatory but worries others may not be as lucky. Many will have to leave the island to pursue their careers. "There is no way that I can stay," dos Santos says. Her son, who was born in Puerto Rico, once dreamed of working at Arecibo.

The exodus is likely to contribute to Puerto Rico's severe brain drain, and will deplete the pool of mentors for student research programs and hands-on training like ENCANTO, a program for undergraduates from underrepresented groups. "The pipeline of Hispanic scientists that came through the observatory is unmatched," says Robert Kerr, a physicist who was director of the observatory for 6 years. "I don't know how exactly we replace that." This year ENCANTO was reduced from 2 years to one; its future is uncertain.

NSF says it's still open to proposals for scientists to visit Arecibo and use the instruments that remain, such as the 12-meter telescope. And for the site of the big scope, scientists including Brisset recently proposed a smaller telescope. But former Arecibo Director Mike Nolan, now an astronomer at the University of Arizona, says the area around the giant telescope is "going to be just a hole in the ground. … The jungle will come back."

For now, researchers are scrambling to get their last bits of data, archive them, and move instruments before they have to leave. "We have 1 month to get the data," physicist Anish Damodaran says of solar observations he made with the 12-meter telescope. Kerr has about $1 million worth of instruments in the optical lab that he must move back to the United States.

NSF is banking on using the site's visitor center and facilities in what will be called the Arecibo Center for STEM Education and Research (ACSER). Officials say they hope the new center will match what the observatory did for education in Puerto Rico by offering student field trips, summer programs, and exhibits. "We want to embrace more of the scientific significance of the area," such as its ecology, says Monya Ruffin, a deputy division director at NSF. The agency expects to award funds to one proposal by the end of the year.

Ubaldo Córdova Figueroa, a professor at the University of Puerto Rico at Mayagüez, has submitted a proposal with more than 30 partners. He envisions two programs, one focused on education and one on research that includes ecology, such as genetic studies of endemic Puerto Rican species. His pitch calls for science camps for students, trainings for teachers, and 3D printing workshops at the former observatory's machine shop.

But Nolan wonders how effective such programs can be if no astronomers or physicists are on site and students can't see science being done. And Cordova Figueroa worries that even if he wins the award, ACSER's funding, totaling $5 million over 5 years, may not be enough for the programs he envisions.

Nine months after the announcement of the transition, Arecibo's future remains murky. On 14 August, "we are going to shut off the lights and put the locks on the doors," Vazquez says. "We don't know what will happen after that." ■

The 2020 collapse of Arecibo's giant radio telescope has upended some scientists' careers.

PHOTO: C. LÓPEZ LLOREDA/*SCIENCE*

PALEOCLIMATE

Lava outburst may have led to Snowball Earth

Eruption dates suggest chemical weathering of rocks triggered plunging temperatures

By **Maya Wei-Haas**

About 717 million years ago, a climate catastrophe struck the planet, as temperatures plunged and glaciers enveloped the globe. The cause of this "Snowball Earth" episode has been mysterious, but it took place around the same time as a massive outburst of volcanism. Many researchers thought there might be a connection. But the timing was uncertain.

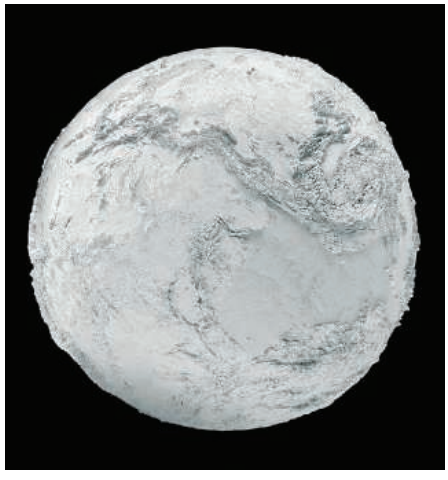
Now, more precise dates, reported last month in *Earth and Planetary Science Letters* (*EPSL*) and in November 2022 in *Science Advances*, show the eruptions preceded the Snowball Earth event by 1 million to 2 million years. The lag points to a particular way the fire could have triggered the ice: through a chemical alteration of the fresh volcanic rocks known as weathering, which sucks carbon dioxide (CO_2) from the atmosphere, turning down the planetary thermostat. The studies highlight the power of weathering as a key driver behind shifts in Earth's climate, and how components of the planet as disparate as rocks and the atmosphere are inextricably linked, says *EPSL* study co-author Galen Halverson, a sedimentary geologist at McGill University. "Nothing can be understood in isolation."

Geoscientists debating the cause of the so-called Sturtian glaciation, which lasted 57 million years, have pointed to a number of possibilities—meteorite strikes, biologic activity, shifts in Earth's orbit, and more. But recent studies have zeroed in on one of the largest volcanic outbursts ever, preserved today across northern Canada in what's called the Franklin large igneous province (LIP). The eruptions spewed lava across an area at least the size of Argentina—and perhaps bigger than China.

Volcanism can trigger cooling in two main ways. In one, eruptions release sulfur-rich gases, which form aerosols that block sunlight and cool the planet. A section of the Franklin lava likely even burst through rocks full of sulfur-rich minerals that could have supercharged the plumes. The other mechanism is weathering. Lava rocks are particularly susceptible to the reactions, in which CO_2 in rainwater reacts with the rocks, ultimately forming minerals that precipitate in the ocean.

"The difference between all these proposals is the timescales on which they happen," says Judy Pu, lead author of the *Science Advances* study at the University of California, Santa Barbara. Sulfur aerosols, for example, only linger in the atmosphere for months to years. Weathering, however, is much slower, requiring 1 million to 2 million years before the cooling effects peak.

Yet until recently, the estimated timing for the Franklin LIP spanned more than 10 million years, making it impossible to discern which of the two mechanisms could have been the trigger. Improvements in dating—and the two teams' sheer determination to recover rare minerals to date—are helping finally pin down the timing. "Everything starts with good ages," says Ashley Gumsley, a geologist specializing in LIPs at the University of Silesia in Katowice who was not part of the study teams.



Some 717 million years ago, glaciers covered the globe. Earlier eruptions may have triggered the event.

By measuring the ratios of trace amounts of uranium and lead trapped in tiny crystals of the mineral zircon—and knowing how fast uranium decays to lead—the teams discovered that the Franklin LIP formed in just 2 million years or so, much faster than most previous estimates. And the primary pulse of volcanism happened 1 million to 2 million years before Snowball Earth, which has been dated through analysis of rocks scoured up by glaciers and eventually deposited in the ocean.

That's exactly the sort of time frame required for weathering to cause the cooling. "They're just stunning the extent to which [the dates] all line up," Halverson says. Additional analysis by Pu and her colleagues suggest the Franklin LIP formed a broad volcanic highland that would have been battered by wind and water, speeding up the weathering.

Pu says other factors may have dialed up the weathering enough to cause a Snowball Earth event. For starters, all of Earth's landmasses at the time were located near the equator, where temperatures were warm and rain frequently pelted the surface. The floods of lava also emerged during the breakup of the Rodinia supercontinent, which exposed fresh rocky surfaces to weathering and may have already begun to cool the global climate. And Gumsley says other large eruptions probably took place at this time in Siberia, China, Africa, and Antarctica, adding to the overall weathering effect.

Case closed? Not for Harvard University geologist emeritus Paul Hoffman, a co-author on the *Science Advances* study who led much of the early work on Snowball Earth. Even if the timing of the Franklin LIP has improved, that of the global glaciation remains slightly uncertain, he says. Forming the ice-scoured rock that marks the start of the episode requires the flow of thick ice at sea level, a process that may not have started until several hundred thousand years after the oceans froze over.

Other Snowball periods remain mysterious. Some 650 million years ago, not long after the Sturtian glaciation, the planet plunged into another deep freeze—but no volcanic outburst preceded it, notes Linda Sohl, a paleoclimatologist at Columbia University who was not part of the study teams. Still, Halverson says weathering may have been a key factor then, too, because all the scouring by glaciers during the earlier Snowball would have exposed fresh rock.

For the Sturtian glaciation at least, the case for a weathering trigger has grown, Sohl says. "In the very least, a revisit with weathering models seems in order," she says. "The snowball glaciations are such fascinating events that are going to keep all of us quite busy for years to come." ■

Maya Wei-Haas is a science journalist in Washington, D.C.

GUINEA PIGBOTS

Doing research with human subjects is costly and cumbersome. Can AI chatbots replace them?

By **Matthew Hutson**; *Illustrations by* **Ashley Mastin**

For Kurt Gray, a social psychologist at the University of North Carolina at Chapel Hill, conducting experiments comes with certain chores. Before embarking on any study, his lab must get ethical approval from an institutional review board, which can take weeks or months. Then his team has to recruit online participants—easier than bringing people into the lab, but Gray says the online subjects are often distracted or lazy. Then the researchers spend hours cleaning the data. But earlier this year, Gray accidentally saw an alternative way to do things.

He was working with computer scientists at the Allen Institute for Artificial Intelligence to see whether they could develop an AI system that made moral judgments like humans. But first they figured they'd see if a system from the startup OpenAI could already do the job. The team asked GPT-3.5, which produces eerily humanlike text, to judge the ethics of 464 scenarios, previously appraised by human subjects, on a scale from −4 (unethical) to 4 (ethical)—scenarios such as selling your house to fund a program for the needy or having an affair with your best friend's spouse. The system's answers, it turned out, were nearly identical to human responses, with a correlation coefficient of 0.95.

"I was like, 'Whoa, we need to back up, because this is crazy,'" Gray says. "If you can just ask GPT to make these judgments, and they align, well, why don't you just ask GPT instead of asking people, at least sometimes?" The results were published this month in *Trends in Cognitive Science* in an article titled "Can AI Language Models Replace Human Participants?"

Generative language models, as these AI systems are known, have taken the world by storm. Perhaps the best known is OpenAI's series of GPT models, which power the ChatGPT chatbot. But other major tech companies, including Google and Meta, are plowing resources into their own models. After being trained on massive amounts of text from books and web pages, these models have an uncanny ability to mimic verbal human behavior. They have already found use in writing computer code, summarizing legal documents, and powering chatbots that tutor students or conduct therapy.

Now, researchers are considering AI's ability to impersonate human subjects in fields such as psychology, political science, economics, and market research. No one is yet suggesting that chatbots can completely replace humans in behavioral studies. But they may act as convenient stand-ins in pilot studies and for designing experiments, saving time and money. Language models might also help with experiments that would be too impractical, unethical, or even dangerous to run with people. "It's a really interesting time," says Ayelet Israeli, a marketing professor at Harvard Business School who believes the models' impact on behavioral research could amount to a "revolution." "Some of these results are just astonishing."

IN HIS ETHICS study, Gray was using GPT-3.5 as a sort of collective everyman, in the hopes of soliciting an average human response. But such models can also be used to populate panels with strikingly diverse participants, because they can be prompted to play anyone: A model contains multitudes. Last year, researchers at Brigham Young University (BYU) created what they call "silicon samples," simulations of human samples. In one study, they fed GPT-3 information about an adopted guise, including age, gender, race, education level, and political affiliation. When the researchers left one of these variables out and asked the model to fill it in, its answers closely matched those from a survey of voters. The researchers also found the model spat out political speech that matched its assigned party affiliation. Lisa Argyle, a BYU political psychologist and co-author on the study, wants to use the virtual participants to test questions for online surveys, identifying those that are most likely to be revealing. That could make actual surveys more efficient. "This is important because survey samples are growing both more expensive and less representative," she says.

Language models can also adopt personality archetypes. In a study led by Hang Jiang, a computer scientist at the Massachusetts Institute of Technology (MIT), researchers had GPT-3.5 assume hundreds of personas by prompting it to behave with different combinations of personality traits—for example, introverted, antagonistic, conscientious, neurotic, and closed to experience. For each persona, they had the model complete a standard personality test and write an 800-word childhood story that was then analyzed for psycholinguistic features associated with personality traits. The models dutifully manifested their assigned personalities both on the test and in the stories. Jiang says such models might allow researchers to test, say, how well people with different personalities would perform in various jobs.

Market researchers are already finding value in the models. In a recent study, Israeli and colleagues found that GPT-3.5 seemed to display realistic consumer behavior. When asked whether it would buy a laptop at various prices, it was less sensitive to price when told its income was $120,000 versus $50,000. It preferred whatever toothpaste brand it had bought previously, and it would pay less for yogurt if it already had a lot at home. It also said it would pay realistic premiums for certain product attributes, such as toothpaste with fluoride and deodorant without aluminum.

The model didn't always give the same answers but instead offered a range of responses about its preferences and willingness to pay. Israeli and her colleagues aggregated its many responses, building a virtual customer survey for these token products for a fraction of the time and money it would have taken in the real world. Language model training data are biased toward Western, affluent people, so the consumer survey might be similarly skewed. But Israeli imagines prompting the AI to impersonate a range of consumers—or to zoom in on a particular demographic—in order to create a more representative study of a product's appeal or potential.

One market research company is already putting language models to work. The startup Synthetic Users has set up a service using OpenAI models in which clients—including Google, IBM, and Apple—can describe a type of person they want to survey, and ask them questions about their needs, desires, and feelings about a product, such as a new website or a wearable. The company's system generates synthetic interviews that co-founder Kwame Ferreira says are "infinitely richer" and more useful than the "bland" feedback companies get when they survey real people.

CHATBOTS CAN ALSO be pitted against one another to study more complex human interactions. Last year, researchers at Stanford University and Google developed "social simulacra" for studying user behavior on platforms such as Facebook and Reddit. The researchers populated a platform they called SimReddit with the equivalent of 1000 different users, by repeatedly prompting GPT-3 with a user identity, a community topic, community rules, and previous posts to the forum. Humans had a hard time distinguishing the resulting discussions from real ones, and platform designers found the tool useful for creating rules or moderation practices.

This year, the researchers built a more immersive simulation populated with what they call "generative agents." The characters were given the ability to remember experiences, reflect on them, and generate and execute plans. Organized behavior emerged: the researchers gave one agent the idea to throw a Valentine's Day party, and over 2 days all the agents in town coordinated to throw one. Joon Sung Park, a Stanford computer science graduate stu-

dent who led both projects, says the virtual world could be used to study the effect of economic policies over time before imposing them on real people.

Economists and psychologists have used agent-based models for many years, programming both the agents and the rules of engagement. But the simulations tend to be simple and depend on hand-coded theoretical assumptions. John Horton, an economist at the MIT Sloan School of Management who has done related work, says agents based on language models are more realistic. He imagines simulating thousands of job seekers and hiring managers to test labor market regulations. "That would be pretty wild," he says.

> "It is plausible that we will have a system within a few years that can just be placed into any experiment and will produce behavior indistinguishable from human behavior."
>
> **Marcel Binz**, Max Planck Institute for Biological Cybernetics

DESPITE ALL their apparent capabilities, the language models are by no means perfect human mirrors. They display several classic human biases but not others. For instance, one recent study of GPT-3.5 found that, like humans, it tends to overestimate how widespread its opinions are in the general population, a bias known as the false consensus effect. But unlike humans, the model showed little hesitation in taking risks and tempting fate. Marcel Binz, a cognitive scientist at the Max Planck Institute for Biological Cybernetics, says AI might need to physically interact with the world to exactly mimic human participants; it's hard to learn all the nuances of intelligent behavior just through passive reading. But he thinks AI will progress quickly in any case. "It is plausible that we will have a system within a few years that can just be placed into any experiment and will produce behavior indistinguishable from human behavior."

A critical question is whether language models will not just reproduce existing findings, but generalize and predict new ones. When models appear to match published psychology studies, they could be regurgitating training data in response to memorized questions. As a result, many researchers are taking pains to phrase questions in novel ways.

Another lingering issue is whether models reflect what people would actually do or just what they say they'd do. People often lie to researchers—and even themselves. Synthetic Users co-founder Hugo Alves suspects the models state true preferences, because they're trained partially on the nakedly honest material contained in anonymous discussion forums. "I've asked things in parenting forums that I wouldn't ask a friend," he says.

Horton worries that the unguarded responses may not last, as OpenAI and others guide their models to be safer and less offensive. "The push to make these models more aligned and not say bad stuff is kind of contrary to social science," he says. "Real people aren't nice all the time. Real people say racist, sexist stuff."

FOR NOW, synthetic participants are most useful for piloting experiments, researchers say. If a model gives unexpected answers to survey questions or doesn't respond at all, Argyle says, your questions may be hard to understand and need rewriting. Israeli says you can design a survey with 1000 questions and use language models to narrow them down to those most likely to correlate with an outcome of interest. Similarly, in economics experiments, Horton says you could run 1 million bargaining scenarios with a model to identify the factors that most affect behavior—before launching the study with people. "The simulation kind of gave you a map," he says.

You could also run experiments you wouldn't ever want to do with people. The 1963 Milgram experiment, in which participants obeyed orders to deliver what they thought were increasingly strong electric shocks to an unseen second set of subjects, probably wouldn't pass ethical review today. But it was easy enough for Gati Aher, an undergraduate computer science student at Olin College of Engineering, to replicate the infamous study with GPT-3. She and her colleagues found that, like the people in the original experiment, the model didn't begin to let off the button until 300 volts.

Aher thinks the models could provide guidance in other sensitive arenas that are difficult to study, for example, what to say to a person who is suicidal. Gray says researchers could study ostracism, or the role of negative feedback on self-esteem. Or, he says, they could study dehumanization of the kind seen in the Vietnam War's Mỹ Lai massacre by describing the situation and asking a model what it would do—provided the models aren't too sanitized.

Argyle says she doesn't know of anyone yet who has replaced human participants with language models. "To be honest, this is all still pretty much a hypothetical," she says. "First we have to demonstrate the language models can do the work." But Horton believes the shift is inevitable. It reminds him of a similar transformation a decade ago, when many social science experiments moved from in-person to online surveys. "People were like, 'How can you run experiments online? Who are these people?' And now it's like, 'Oh, yeah, of course you do that.'"

Chatbots might already be infiltrating online surveys—but among subjects rather than researchers. A recent study asked crowdworkers to summarize some text and found that at least one-third were likely using ChatGPT. Gray says, half-jokingly, "If online participants are already using GPT, we might as well just ask GPT itself." ■

Matthew Hutson is a journalist in New York City.

ILLUSTRATION: A. MASTIN/*SCIENCE*

INSIGHTS

PERSPECTIVES

PLANT SCIENCE

Genetic editing of wood for sustainability

Trees engineered to have less lignin could make paper production less polluting

By **Vânia G. Zuin Zeidler**

Lignin, a polymer formed by phenylpropanoid units, is responsible for the rigidity and resistance of the lignocellulosic cells in wood (*1*). In conventional pulp production, lignin must be cleaved and dissolved under alkaline conditions or first sulfonated to make it soluble so that fiber separation can take place. Delignification processes are reagent and energy intensive, leading to costly chemical recovery (*2*). Pulp treatment methods to remove wood extractives such as lignin have been developed, but they are not yet economically viable at an industrial scale (*3*). On page 216 of this issue, Sulis *et al.* (*4*) present a multiplex CRISPR genome editing strategy to modify lignin biosynthesis genes and reduce the lignin content of *Populus trichocarpa*, a species of poplar. This approach could provide a solution to a key operational constraint in the paper and pulp industry.

Around 50 million tons of lignin are generated per year as a by-product in the paper and pulp industry (*5*). In 2021, the industry produced greenhouse gases with a global warming potential equivalent to 190 million tons of carbon dioxide, a figure that is projected to increase in the next decades (*6*). In addition, almost 1.7 billion tons of feedstocks and reagents were used to produce around 820 million metric tons of final chemical products in 2015, while also producing around the same quantity of chemical waste (*7*). These statistics reveal resource and process inefficiencies that have resulted in substantial chemical releases to environmental compartments, including the atmosphere, soil, water, and animal and plant life (*8*).

Most commercially valuable woods are genetically complex, which makes it difficult to determine which genes contribute to the complex phenotypes and associate them with causal mechanisms, such as lignin biosynthesis (*9*). Therefore, recently developed bioengineering approaches such as multiplex CRISPR genome editing that allow multiple genes to be edited in a single effort—a goal that was more difficult to achieve with previous approaches—are likely to become important for developing a new generation of woody biorefinery crops (*10*, *11*). Indeed, multiplex genome editing of other plants has already been used to modify traits such as herbicide resistance, hormone biosynthesis, and metabolic function (*12*).

Sulis *et al.* identified target genes for their study with a previously established computational predictive model for monolignol biosynthesis in poplar trees (*13*). The oxidative polymerization of three major monolignols is the predominant method of lignin formation in flowering plants such as poplar (*14*). The aim of the model was to identify genetic modifications that would result in reduced lignin content and increased syringyl-to-guaiacyl and carbohydrate-to-lignin ratios. Seven strategies involving the modification of three to six genes each and with the best predicted fiber trait improvements were selected for multiplex editing in poplar wood.

A multiplex CRISPR construct was assembled for each strategy and delivered to poplar trees using *Agrobacterium tumefaciens*, generating 174 independent lines of edited poplar wood. The wood of a 6-month-old wild-type poplar tree contained 20.9% lignin and had a carbohydrate-to-lignin ratio of 3.0, whereas in edited wood, lignin content was reduced by up to 49.1%, and the carbohydrate-to-lignin ratio was increased by up to 228%. However, it should be noted that several of the edited lines with substantial reductions in lignin content also had reduced tree growth, demanding additional studies to establish the impact of such a change on

Institute of Sustainable Chemistry Faculty, School of Sustainability, Leuphana University, Lüneburg, Germany. Email: vania.zuin@leuphana.de

Wood from *Populus trichocarpa*, a variety of poplar tree, is used to manufacture paper, a process that can be highly polluting.

the pulp and paper industry. These studies should also assess wood density and the performance of edited trees in actual plantations, with a higher number of replicates per line. In Europe and North America, poplar has become an important source of wood for pulp and paper, but it is not the only tree used for this purpose. In the future, it would be interesting to apply the approach taken by Sulis *et al.* to other wood varieties.

Sulis *et al.* modeled the techno-economic impacts of CRISPR-edited wood on kraft pulping and found that yield and production efficiency were increased, resulting in reduced need for pulping chemicals. Additionally, they estimated that pulp production using edited wood would reduce the global warming potential of the process by up to 20%. These changes would be a step toward a more sustainable fiber production system, which would align with the United Nations Sustainable Development Goals (*15*).

In the European Union, engineered trees such as those produced by Sulis *et al.* are considered genetically modified organisms and must comply with the directive 2001/18/EC. This means that the European regulations would limit, or even forbid, the market introduction of gene-edited trees. However, some suggest that there is no reason to differentiate gene-edited trees with small deletions or insertions from the ones produced by conventional breeding or mutagenesis and that, to face the current challenges related to environmental impacts, biosafety regulations should be based on the specific risks and benefits of each product (*9*). ■

REFERENCES AND NOTES

1. I. F. Demuner, J. L. Colodette, A. J. Demuner, C. M. Jardim, *BioResources* **14**, 7543 (2019).
2. L. Chen *et al.*, *Sci. Adv.* **3**, e1701735 (2017).
3. M. Lehr, M. Miltner, A. Friedl, *SN Appl. Sci.* **3**, 886 (2021).
4. D. B. Sulis *et al.*, *Science* **381**, 216 (2023).
5. I. F. Demuner *et al.*, *J. Clean. Prod.* **317**, 128286 (2021).
6. International Energy Agency, Pulp and Paper (IEA, 2022); https://www.iea.org/reports/pulp-and-paper.
7. United Nations Environment Programme, *Green and Sustainable Chemistry: Framework Manual* (UNEP, 2020).
8. M. C. Monte, E. Fuente, A. Blanco, C. Negro, *Waste Manag.* **29**, 293 (2009).
9. C. Anders, L. Hoengenaert, W. Boerjan, *Curr. Opin. Plant Biol.* **71**, 102329 (2023).
10. A. A. Myburg, S. G. Hussey, J. P. Wang, N. R. Street, E. Mizrachi, *Front. Plant Sci.* **10**, 775 (2019).
11. V. G. Zuin, *Curr. Opin. Green Sustain. Chem.* **2**, 40 (2016).
12. V. A. Najera, R. M. Twyman, P. Christou, C. Zhu, *Curr. Opin. Biotechnol.* **59**, 93 (2019).
13. M. L. Matthews, J. P. Wang, R. Sederoff, V. L. Chiang, C. M. Williams, *Comput. Struct. Biotechnol. J.* **19**, 168 (2020).
14. J. P. Wang *et al.*, *Nat. Commun.* **9**, 1579 (2018).
15. K. Kümmerer, J. H. Clark, V. G. Zuin, *Science* **367**, 369 (2020).

ACKNOWLEDGMENTS

The author thanks J. Coury for helpful comments.

10.1126/science.adi8186

EPIGENETICS

Metabolic control of transcription

Light alters histone methylation in plants via nuclear α-ketoglutarate dehydrogenase

By **Eric B. Taylor**

Mechanisms that link environmental stimuli to organismal transcriptional responses facilitate adaptation and survival. Because it signals energy availability through photosynthesis, light is among the most important environmental signals for plants. On page 179 of this issue, Huang *et al.* (*1*) demonstrate that, in the plant *Arabidopsis thaliana*, light exposure stimulates chromatin-local depletion of the tricarboxylic acid (TCA) cycle metabolite α-ketoglutarate to orchestrate global gene expression by impairing demethylation of nucleosomal histone tails. These findings provide a molecular mechanism for how the *Arabidopsis* genome senses light. They also demonstrate how subcellular metabolism can be spatially reorganized to regulate chromatin structure and transcription.

Methylation of histone-tail lysine (K) residues is a conserved mechanism for both activating and repressing gene expression (*2, 3*). Nuclear DNA wraps around histones to form chromatin, and modification of histone-tail lysines alters chromatin structure and modulates interaction with transcription control machinery. Methylation (me) patterns are diverse, including mono (me1)-, di (me2)-, and tri (me3)–methylation, with lysines on histone 3 (H3) known to be especially important for regulating gene expression.

Jumonji C domain–containing histone demethylases (JMJs) are a major class of histone demethylases that use α-ketoglutarate as a rate-limiting substrate to remove methyl groups from histone lysines by oxidation (*4–6*). In the TCA cycle, α-ketoglutarate is produced by isocitrate dehydrogenase (IDH) and catabolized by α-ketoglutarate dehydrogenase (KGDH) to succinyl–coenzyme A. Using subcellular fractionation, light microscopy, and immunoprecipitation, Huang *et al.* showed that after light exposure, a small fraction of KGDH translocates into the nucleus and physically interacts with JMJs (see the figure).

KGDH comprises three subunits: E1, E2, and E3. Huang *et al.* found that compared with wild-type cells, whole cellular genetic disruption of either E1 or E2 subunits decreased the abundance of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 methylation marks. This decrease was accompanied by late flowering of the plant and accumulation of antioxidant anthocyanin pigments in seedlings. RNA sequencing revealed broad changes in expression of genes related to abiotic and biotic stress responses, which is consistent with perturbed sensing of the environment.

Because of its central role in cellular metabolism, genetic ablation of KGDH could lead to distal, nonspecific stress responses that affect histone methylation and gene expression. Thus, a key challenge faced by Huang *et al.* was to parse global effects of metabolic stress from those specific to loss of nuclear α-ketoglutarate. Through experiments in which the genes encoding KGDH E1 and E2 subunits were deleted and then added back with or without mutations, they determined that regulation of gene expression by KGDH depends on nuclear localization conferred by the nuclear localization sequence in the E2 subunit.

Among genes that might explain the observed late flowering and anthocyanin accumulation, the central repressor of flowering *FLOWERING LOCUS C* (*FLC*) and master regulator of anthocyanin biosynthesis *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*PAP1*) genes were both derepressed by genetic KGDH disruption. The authors performed tests of causality by mutating either *FLC* or *PAP1* in the presence of KGDH disruption. These secondary mutations largely rescued late flowering and anthocyanin accumulation, respectively. Through this result, Huang *et al.* show that, although many metabolic signaling events clearly evolved as feedback mechanisms to tightly maintain metabolic homeostasis (*7, 8*), metabolic signaling out-

Department of Molecular Physiology and Biophysics and Fraternal Orders of Eagle Diabetes Research Center, Holden Comprehensive Cancer Center, Abboud Cardiovascular Research Center, Pappajohn Biomedical Discovery Institute, Carver College of Medicine, University of Iowa, Iowa City, IA, USA. Email: eric-taylor@uiowa.edu

Wood from *Populus trichocarpa*, a variety of poplar tree, is used to manufacture paper, a process that can be highly polluting.

the pulp and paper industry. These studies should also assess wood density and the performance of edited trees in actual plantations, with a higher number of replicates per line. In Europe and North America, poplar has become an important source of wood for pulp and paper, but it is not the only tree used for this purpose. In the future, it would be interesting to apply the approach taken by Sulis *et al.* to other wood varieties.

Sulis *et al.* modeled the techno-economic impacts of CRISPR-edited wood on kraft pulping and found that yield and production efficiency were increased, resulting in reduced need for pulping chemicals. Additionally, they estimated that pulp production using edited wood would reduce the global warming potential of the process by up to 20%. These changes would be a step toward a more sustainable fiber production system, which would align with the United Nations Sustainable Development Goals (*15*).

In the European Union, engineered trees such as those produced by Sulis *et al.* are considered genetically modified organisms and must comply with the directive 2001/18/EC. This means that the European regulations would limit, or even forbid, the market introduction of gene-edited trees. However, some suggest that there is no reason to differentiate gene-edited trees with small deletions or insertions from the ones produced by conventional breeding or mutagenesis and that, to face the current challenges related to environmental impacts, biosafety regulations should be based on the specific risks and benefits of each product (*9*). ■

REFERENCES AND NOTES

1. I. F. Demuner, J. L. Colodette, A. J. Demuner, C. M. Jardim, *BioResources* **14**, 7543 (2019).
2. L. Chen *et al.*, *Sci. Adv.* **3**, e1701735 (2017).
3. M. Lehr, M. Miltner, A. Friedl, *SN Appl. Sci.* **3**, 886 (2021).
4. D. B. Sulis *et al.*, *Science* **381**, 216 (2023).
5. I. F. Demuner *et al.*, *J. Clean. Prod.* **317**, 128286 (2021).
6. International Energy Agency, Pulp and Paper (IEA, 2022); https://www.iea.org/reports/pulp-and-paper.
7. United Nations Environment Programme, *Green and Sustainable Chemistry: Framework Manual* (UNEP, 2020).
8. M. C. Monte, E. Fuente, A. Blanco, C. Negro, *Waste Manag.* **29**, 293 (2009).
9. C. Anders, L. Hoengenaert, W. Boerjan, *Curr. Opin. Plant Biol.* **71**, 102329 (2023).
10. A. A. Myburg, S. G. Hussey, J. P. Wang, N. R. Street, E. Mizrachi, *Front. Plant Sci.* **10**, 775 (2019).
11. V. G. Zuin, *Curr. Opin. Green Sustain. Chem.* **2**, 40 (2016).
12. V. A. Najera, R. M. Twyman, P. Christou, C. Zhu, *Curr. Opin. Biotechnol.* **59**, 93 (2019).
13. M. L. Matthews, J. P. Wang, R. Sederoff, V. L. Chiang, C. M. Williams, *Comput. Struct. Biotechnol. J.* **19**, 168 (2020).
14. J. P. Wang *et al.*, *Nat. Commun.* **9**, 1579 (2018).
15. K. Kümmerer, J. H. Clark, V. G. Zuin, *Science* **367**, 369 (2020).

ACKNOWLEDGMENTS

The author thanks J. Coury for helpful comments.

10.1126/science.adi8186

EPIGENETICS

Metabolic control of transcription

Light alters histone methylation in plants via nuclear α-ketoglutarate dehydrogenase

By **Eric B. Taylor**

Mechanisms that link environmental stimuli to organismal transcriptional responses facilitate adaptation and survival. Because it signals energy availability through photosynthesis, light is among the most important environmental signals for plants. On page 179 of this issue, Huang *et al.* (*1*) demonstrate that, in the plant *Arabidopsis thaliana*, light exposure stimulates chromatin-local depletion of the tricarboxylic acid (TCA) cycle metabolite α-ketoglutarate to orchestrate global gene expression by impairing demethylation of nucleosomal histone tails. These findings provide a molecular mechanism for how the *Arabidopsis* genome senses light. They also demonstrate how subcellular metabolism can be spatially reorganized to regulate chromatin structure and transcription.

Methylation of histone-tail lysine (K) residues is a conserved mechanism for both activating and repressing gene expression (*2, 3*). Nuclear DNA wraps around histones to form chromatin, and modification of histone-tail lysines alters chromatin structure and modulates interaction with transcription control machinery. Methylation (me) patterns are diverse, including mono (me1)-, di (me2)-, and tri (me3)–methylation, with lysines on histone 3 (H3) known to be especially important for regulating gene expression.

Jumonji C domain–containing histone demethylases (JMJs) are a major class of histone demethylases that use α-ketoglutarate as a rate-limiting substrate to remove methyl groups from histone lysines by oxidation (*4–6*). In the TCA cycle, α-ketoglutarate is produced by isocitrate dehydrogenase (IDH) and catabolized by α-ketoglutarate dehydrogenase (KGDH) to succinyl–coenzyme A. Using subcellular fractionation, light microscopy, and immunoprecipitation, Huang *et al.* showed that after light exposure, a small fraction of KGDH translocates into the nucleus and physically interacts with JMJs (see the figure).

KGDH comprises three subunits: E1, E2, and E3. Huang *et al.* found that compared with wild-type cells, whole cellular genetic disruption of either E1 or E2 subunits decreased the abundance of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 methylation marks. This decrease was accompanied by late flowering of the plant and accumulation of antioxidant anthocyanin pigments in seedlings. RNA sequencing revealed broad changes in expression of genes related to abiotic and biotic stress responses, which is consistent with perturbed sensing of the environment.

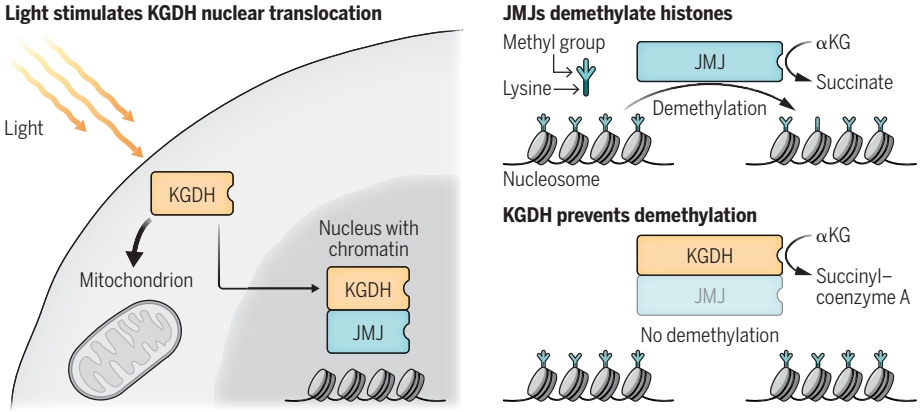
Because of its central role in cellular metabolism, genetic ablation of KGDH could lead to distal, nonspecific stress responses that affect histone methylation and gene expression. Thus, a key challenge faced by Huang *et al.* was to parse global effects of metabolic stress from those specific to loss of nuclear α-ketoglutarate. Through experiments in which the genes encoding KGDH E1 and E2 subunits were deleted and then added back with or without mutations, they determined that regulation of gene expression by KGDH depends on nuclear localization conferred by the nuclear localization sequence in the E2 subunit.

Among genes that might explain the observed late flowering and anthocyanin accumulation, the central repressor of flowering *FLOWERING LOCUS C* (*FLC*) and master regulator of anthocyanin biosynthesis *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*PAP1*) genes were both derepressed by genetic KGDH disruption. The authors performed tests of causality by mutating either *FLC* or *PAP1* in the presence of KGDH disruption. These secondary mutations largely rescued late flowering and anthocyanin accumulation, respectively. Through this result, Huang *et al.* show that, although many metabolic signaling events clearly evolved as feedback mechanisms to tightly maintain metabolic homeostasis (*7, 8*), metabolic signaling out-

Department of Molecular Physiology and Biophysics and Fraternal Orders of Eagle Diabetes Research Center, Holden Comprehensive Cancer Center, Abboud Cardiovascular Research Center, Pappajohn Biomedical Discovery Institute, Carver College of Medicine, University of Iowa, Iowa City, IA, USA. Email: eric-taylor@uiowa.edu

α-Ketoglutarate dehydrogenase prevents histone demethylation

In cells of the plant *Arabidopsis thaliana*, light exposure stimulates the translocation of a small portion of α-ketoglutarate dehydrogenase (KGDH) to the nucleus, where it interacts with Jumonji C domain–containing histone demethylases (JMJs). JMJs remove methyl groups from histone lysines using α-ketoglutarate (αKG) as a substrate. Nuclear KGDH depletes nuclear αKG, preventing histone demethylation and altering transcription.



puts do not necessarily require intact metabolism. Thus, late flowering arose from altered metabolite signaling, not a primary metabolic impairment.

Several JMJ demethylase families are found across eukaryotic species (*9*). Because individual JMJs have different affinities and selectivities for methylated histone lysines, it is possible that regulated, JMJ family member–specific association with KGDH confers locus-specific control. In support of this possibility, through chromatin immunoprecipitation–quantitative polymerase chain reaction (PCR) experiments, Huang *et al.* observed the specific regulation of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 methylation status by KGDH. They also observed selective occupancy of different KGDH-JMJ complexes within the gene bodies, 5′-transcribed regions, or both of different genes.

Several important questions will need to be addressed in future research. The signal transduction mechanism by which light stimulates nuclear KGDH translocation remains undefined. Deciphering the signaling transducers of light exposure could reveal additional inputs from nonlight environmental stimuli that converge on common signaling hubs to increase nuclear KGDH activity. In this case, the precision of transcriptional responses to discreet stimuli could be provided by additional, event-specific signals that modulate the composition of KGDH-JMJ and other regulatory protein complexes. Furthermore, mapping the full light signal transduction pathway could also provide opportunities for future research to test whether KGDH and other signaling mechanisms may be genetically modulated to enhance plant growth in low light.

Although Huang *et al.* showed that as in *Arabidopsis*, human KGDH and JMJs physically interact, the human E2 KGDH subunit lacks the nuclear localization sequence present in *Arabidopsis* that the study demonstrated is required for nuclear KGDH translocation and repression of histone demethylation. Thus, future work will be required to test whether this phenomenon persists in mammals, with KGDH nuclear translocation potentially regulated by a different mechanism. This is intriguing, because although aberrant mitochondrial metabolism in cancer systems has been shown to increase levels of succinate and 2-hyroxyglutarate, which inhibit JMJ activity (*10, 11*), a mechanism for activating mammalian JMJs has not been identified. ■

REFERENCES AND NOTES

1. F. Huang *et al.*, *Science* **381**, eadf8822 (2023).
2. J. C. Black, C. Van Rechem, J. R. Whetstine, *Mol. Cell* **48**, 491 (2012).
3. A. Jambhekar, A. Dhall, Y. Shi, *Nat. Rev. Mol. Cell Biol.* **20**, 625 (2019).
4. X. Chen, Y. Hu, D.-X. Zhou, *Biochim. Biophys. Acta. Gene Regul. Mech.* **1809**, 421 (2011).
5. S. M. Kooistra, K. Helin, *Nat. Rev. Mol. Cell Biol.* **13**, 297 (2012).
6. S. Markolovic *et al.*, *Curr. Opin. Struct. Biol.* **41**, 62 Epub 20160614 (2016).
7. K. G. Hicks *et al.*, *Science* **379**, 996 (2023).
8. Y. Liu, K. Birsoy, *Mol. Cell* **83**, 877 (2023).
9. N. Mosammaparast, Y. Shi, *Annu. Rev. Biochem.* **79**, 155 (2010).
10. S. L. Campbell, K. E. Wellen, *Mol. Cell* **71**, 398 (2018).
11. L. W. S. Finley, *Cell* **186**, 1670 (2023).

ACKNOWLEDGMENTS

E.B.T. is supported by NIH grant DK104998 and a Distinguished Scholars Award from the University of Iowa.

10.1126/science.adi7577

MICROBIOLOGY

Coupling proteins, with deadly consequences

Phage protein E shuts down cell wall biosynthesis to kill bacteria

By **Jochen Balbach**[1,2] *and* **Milton T. Stubbs**[2,3]

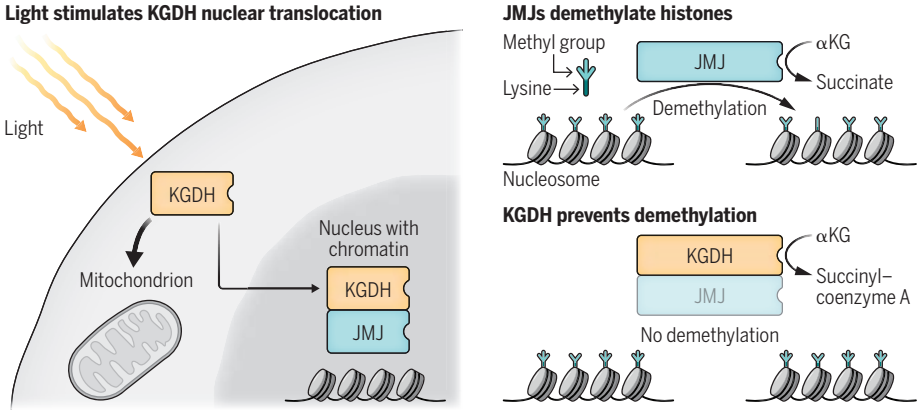
To release infectious progeny, many bacteriophages induce lysis of the host bacterium. One intriguing mechanism involves the single-gene lysis (SGL) phage proteins (*1*). In contrast to antimicrobial peptides, most of which interact directly with the bacterial membrane to form membrane pores (*2*), SGL proteins induce the host cell to undergo autolysis. Expression of the 91-residue transmembrane protein E from the archetypal phage ΦX174 is sufficient to induce autolysis in *Escherichia coli* (*3*). On page 180 of this issue, Orta *et al.* (*4*) report the structure of protein E in complex with host cell proteins MraY (phospho-MurNAc-pentapeptide translocase) and SlyD (sensitive to lysis D). Protein E enlists SlyD to inhibit MraY and thus peptidoglycan biosynthesis. Ensuing defects in the cell wall impair resistance to turgor pressure, leading to host cell lysis. Phages engineered to express protein E could serve as an antibacterial therapeutic strategy.

Almost 40 years ago, the ability of protein E to induce bacteriolysis was shown to be dependent on the presence of a host cytoplasmic protein, SlyD (*3*). SlyD exhibits three distinct chaperone activities: The FKBP (FK506 binding protein) domain acts as a peptidyl-prolyl cis-trans isomerase (*5*, *6*), the IF ("insert in flap") chaperone domain stabilizes protein-folding intermediates through donor-strand complementation (*7*), and the C-terminal histidine-rich sequence is implicated in metalloprotein assembly (*8*). However, until now, none of these activities could be

[1]Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. [2]ZIK HALOmem, Charles-Tanford-Proteinzentrum, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. [3]Institut für Biochemie und Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. Email: stubbs@biochemtech.uni-halle.de

α-Ketoglutarate dehydrogenase prevents histone demethylation

In cells of the plant *Arabidopsis thaliana*, light exposure stimulates the translocation of a small portion of α-ketoglutarate dehydrogenase (KGDH) to the nucleus, where it interacts with Jumonji C domain–containing histone demethylases (JMJs). JMJs remove methyl groups from histone lysines using α-ketoglutarate (αKG) as a substrate. Nuclear KGDH depletes nuclear αKG, preventing histone demethylation and altering transcription.



puts do not necessarily require intact metabolism. Thus, late flowering arose from altered metabolite signaling, not a primary metabolic impairment.

Several JMJ demethylase families are found across eukaryotic species (*9*). Because individual JMJs have different affinities and selectivities for methylated histone lysines, it is possible that regulated, JMJ family member–specific association with KGDH confers locus-specific control. In support of this possibility, through chromatin immunoprecipitation–quantitative polymerase chain reaction (PCR) experiments, Huang *et al.* observed the specific regulation of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 methylation status by KGDH. They also observed selective occupancy of different KGDH-JMJ complexes within the gene bodies, 5′-transcribed regions, or both of different genes.

Several important questions will need to be addressed in future research. The signal transduction mechanism by which light stimulates nuclear KGDH translocation remains undefined. Deciphering the signaling transducers of light exposure could reveal additional inputs from nonlight environmental stimuli that converge on common signaling hubs to increase nuclear KGDH activity. In this case, the precision of transcriptional responses to discreet stimuli could be provided by additional, event-specific signals that modulate the composition of KGDH-JMJ and other regulatory protein complexes. Furthermore, mapping the full light signal transduction pathway could also provide opportunities for future research to test whether KGDH and other signaling mechanisms may be genetically modulated to enhance plant growth in low light.

Although Huang *et al.* showed that as in *Arabidopsis*, human KGDH and JMJs physically interact, the human E2 KGDH subunit lacks the nuclear localization sequence present in *Arabidopsis* that the study demonstrated is required for nuclear KGDH translocation and repression of histone demethylation. Thus, future work will be required to test whether this phenomenon persists in mammals, with KGDH nuclear translocation potentially regulated by a different mechanism. This is intriguing, because although aberrant mitochondrial metabolism in cancer systems has been shown to increase levels of succinate and 2-hyroxyglutarate, which inhibit JMJ activity (*10, 11*), a mechanism for activating mammalian JMJs has not been identified. ■

REFERENCES AND NOTES

1. F. Huang *et al.*, *Science* **381**, eadf8822 (2023).
2. J. C. Black, C. Van Rechem, J. R. Whetstine, *Mol. Cell* **48**, 491 (2012).
3. A. Jambhekar, A. Dhall, Y. Shi, *Nat. Rev. Mol. Cell Biol.* **20**, 625 (2019).
4. X. Chen, Y. Hu, D.-X. Zhou, *Biochim. Biophys. Acta. Gene Regul. Mech.* **1809**, 421 (2011).
5. S. M. Kooistra, K. Helin, *Nat. Rev. Mol. Cell Biol.* **13**, 297 (2012).
6. S. Markolovic *et al.*, *Curr. Opin. Struct. Biol.* **41**, 62 Epub 20160614 (2016).
7. K. G. Hicks *et al.*, *Science* **379**, 996 (2023).
8. Y. Liu, K. Birsoy, *Mol. Cell* **83**, 877 (2023).
9. N. Mosammaparast, Y. Shi, *Annu. Rev. Biochem.* **79**, 155 (2010).
10. S. L. Campbell, K. E. Wellen, *Mol. Cell* **71**, 398 (2018).
11. L. W. S. Finley, *Cell* **186**, 1670 (2023).

ACKNOWLEDGMENTS

E.B.T. is supported by NIH grant DK104998 and a Distinguished Scholars Award from the University of Iowa.

10.1126/science.adi7577

MICROBIOLOGY

Coupling proteins, with deadly consequences

Phage protein E shuts down cell wall biosynthesis to kill bacteria

By **Jochen Balbach**[1,2] *and* **Milton T. Stubbs**[2,3]

To release infectious progeny, many bacteriophages induce lysis of the host bacterium. One intriguing mechanism involves the single-gene lysis (SGL) phage proteins (*1*). In contrast to antimicrobial peptides, most of which interact directly with the bacterial membrane to form membrane pores (*2*), SGL proteins induce the host cell to undergo autolysis. Expression of the 91-residue transmembrane protein E from the archetypal phage ΦX174 is sufficient to induce autolysis in *Escherichia coli* (*3*). On page 180 of this issue, Orta *et al.* (*4*) report the structure of protein E in complex with host cell proteins MraY (phospho-MurNAc-pentapeptide translocase) and SlyD (sensitive to lysis D). Protein E enlists SlyD to inhibit MraY and thus peptidoglycan biosynthesis. Ensuing defects in the cell wall impair resistance to turgor pressure, leading to host cell lysis. Phages engineered to express protein E could serve as an antibacterial therapeutic strategy.

Almost 40 years ago, the ability of protein E to induce bacteriolysis was shown to be dependent on the presence of a host cytoplasmic protein, SlyD (*3*). SlyD exhibits three distinct chaperone activities: The FKBP (FK506 binding protein) domain acts as a peptidyl-prolyl cis-trans isomerase (*5*, *6*), the IF ("insert in flap") chaperone domain stabilizes protein-folding intermediates through donor-strand complementation (*7*), and the C-terminal histidine-rich sequence is implicated in metalloprotein assembly (*8*). However, until now, none of these activities could be

[1]Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. [2]ZIK HALOmem, Charles-Tanford-Proteinzentrum, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. [3]Institut für Biochemie und Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Halle (Saale), Germany. Email: stubbs@biochemtech.uni-halle.de

correlated with SlyD's role in bacteriolysis.

The 34-residue transmembrane N-terminal domain (NTD) of protein E is necessary, but not sufficient, for inducing lysis (*3*); this requires the presence of the 57–amino acid cytosolic C-terminal domain (CTD), which varies in sequence among ΦX174-related phages. Notably, replacement of the CTD with an autonomously folding domain results in lysis even in the absence of *slyD*. Screening of lysis-inducing protein E mutants in a *slyD*-null background revealed the involvement of a second host protein, MraY (*9*). This dimeric integral membrane enzyme (*10*) catalyzes a crucial step in cell wall biosynthesis, namely formation of the prenylated peptidoglycan precursor lipid I, suggesting that protein E acts by disrupting the integrity of the bacterial cell wall.

Orta *et al.* pull these disparate strands together to elucidate the role of each player in bacteriolysis. Coexpression of *E. coli* MraY with the ΦX174 E protein resulted in a stable complex that copurified with *E. coli* SlyD, which the authors dubbed the YES complex (MraY-E-SlyD). Using single-particle cryo–electron microscopy, the structures of $YES_{\Phi X174}$ and a related complex could be traced to single-residue precision, with the tripartite complex forming a dimer of trimers (see the figure).

The study of Orta *et al.* reveals a surprisingly simple yet elegant principle behind protein E action. The NTD hugs a groove in the surface of one monomer of MraY in a helical conformation, with kinks at conserved prolines in protein E. A comparison with the carbacaprazamycin-inhibited MraY structure (*11*) shows that the kinks allow the peptide inhibitor to block access to both the lipid and the nucleoside binding sites simultaneously, effectively supressing substrate binding. The CTD of protein E, which is required for lysis in wild-type *E. coli*, is held in place by two SlyD molecules. An amphiphilic α helix of protein E traverses the cytoplasmic face of MraY parallel to the membrane, sandwiched between the enzyme's active site and the IF domain of one SlyD molecule. The peptide then proceeds in an extended conformation to bind to the IF′ domain of the second SlyD′ molecule through donor-strand complementation, ending with Pro^{65} occupying the second SlyD′ FKBP′ domain active site. Together with a second protein E′ moiety, both active sites of the MraY dimer are blocked. It is remarkable that there are no direct contacts between MraY and SlyD or between the two chaperone molecules SlyD and SlyD′, with protein E mediating all interactions between the host proteins.

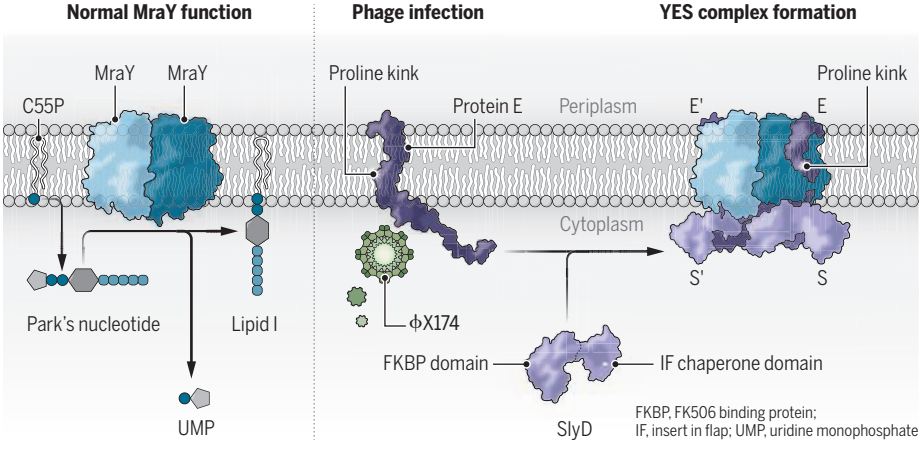
Thus, phage protein E couples two functionally unrelated bacterial proteins, MraY and SlyD. This supresses MraY substrate binding, which inhibits peptidoglycan biosynthesis and thereby initiates host cell lysis. The chaperone function of SlyD is co-opted to stabilize the CTD of protein E, preventing cellular degradation of its (presumably) otherwise disordered C terminus. Why should protein E have evolved to hijack SlyD rather than make use of a folded CTD? The answer lies in the structure of the phage genome: The gene encoding protein E is embedded as an alternative reading frame in the gene of scaffolding protein D. Maintaining the integrity of the protein D gene puts limitations on the genetic space available to that of protein E. Indeed, the necessity to encode multiple functional proteins in a restricted genome is reflected in the diversity found in SGL target proteins (including MurA, which catalyzes the first committed step in peptidoglycan biosynthesis, and MurJ, the conserved lipid II flippase) and accessory proteins (including the chaperone DnaJ) (*1*). Overall, this suggests that lytic functions have arisen spontaneously and opportunistically.

It is not only phages that have co-opted the potential of SlyD. Fusion of one or two copies of the chaperone to the viral envelope proteins HIV gp41 (*12*) and rubella virus E1 (*13*) facilitates high-yield recombinant protein production in *E. coli*. The SlyD fusion prevents aggregation of the viral-antigen protein fragments, maintaining antibody binding competency. This has fostered the commercialization of diagnostic kits for serological quantification of antiviral antibodies in patients. This procedure was used to stabilize the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) receptor binding domain of the Spike protein, which facilitated the rapid development of the double-antigen sandwich assay Elecsys Anti-SARS-CoV-2-S for determining antibody titers in patients after SARS-CoV-2 infection or vaccination (*14*).

More than a century ago, phage pioneer Félix Hubert d'Hérelle recognized the possibilities afforded by phage-induced bacteriolysis in the treatment and prevention of bacterial infections (*15*). Eclipsed after World War II by the advent of antibiotics such as penicillin, research into the therapeutic and prophylactic use of bacteriophages continued in the USSR. The ever-increasing emergence of antibiotic-resistant bacterial strains has reawakened interest in phage therapy. However, bacteriophage use is limited by their narrow strain specificity, so that applications tend to require mixtures of phages. Understanding the molecular mechanism of ΦX174 protein E raises the possibility of designing lytic phages—or equipping nonlytic phages with lytic genes—for the targeted eradication of antibiotic-resistant bacteria. ■

An antibacterial complex

The bacterial enzyme MraY (phospho-MurNAc-pentapeptide translocase) catalyzes a crucial step in peptidoglycan synthesis, namely the formation of lipid I from Park's nucleotide and undecaprenyl phosphate (C55P). However, when *Escherichia coli* are infected with phage ΦX174, the encoded phage protein E couples the host proteins SlyD (sensitive to lysis D) and MraY, forming the YES complex. This inhibits peptidoglycan synthesis and induces bacterial cell lysis.



FKBP, FK506 binding protein; IF, insert in flap; UMP, uridine monophosphate

GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. K. R. Chamakura, R. Young, *Curr. Opin. Microbiol.* **56**, 109 (2020).
2. K. A. Brogden, *Nat. Rev. Microbiol.* **3**, 238 (2005).
3. D. Maratea, K. Young, R. Young, *Gene* **40**, 39 (1985).
4. A. K. Orta *et al.*, *Science* **381**, eadg9091 (2023).
5. W. D. Roof *et al.*, *J. Biol. Chem.* **269**, 2902 (1994).
6. S. Hottenrott *et al.*, *J. Biol. Chem.* **272**, 15697 (1997).
7. C. Löw *et al.*, *J. Mol. Biol.* **398**, 375 (2010).
8. M. J. Lacasse, D. B. Zamble, *Biochemistry* **55**, 1689 (2016).
9. T. G. Bernhardt, W. D. Roof, R. Young, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4297 (2000).
10. B. C. Chung *et al.*, *Science* **341**, 1012 (2013).
11. E. H. Mashalidis *et al.*, *Nat. Commun.* **10**, 2917 (2019).
12. C. Scholz *et al.*, *J. Mol. Biol.* **345**, 1229 (2005).
13. C. Scholz *et al.*, *Biochemistry* **47**, 4276 (2008).
14. P. Muench *et al.*, *J. Clin. Microbiol.* **58**, e01694 (2020).
15. K. E. Kortright *et al.*, *Cell Host Microbe* **25**, 219 (2019).

10.1126/science.adi7571

RETROSPECTIVE

Donald D. Brown (1931–2023)

Father of molecular embryology

By **Steven McKnight**[1] *and* **Allan Spradling**[2]

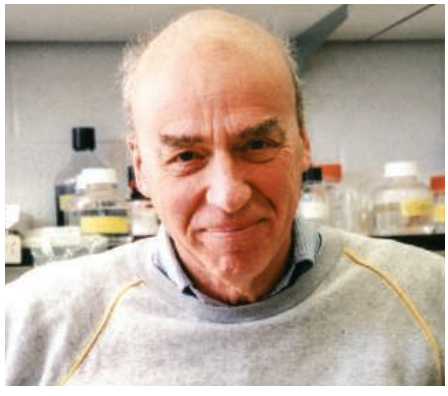
Donald D. ("Don") Brown, groundbreaking biologist who studied isolated genes in test tubes, died on 31 May. He was 91. The technique Brown pioneered, which he called "developmental genetics by gene isolation," allowed him to, for the first time, purify an animal gene and elucidate its structure and regulatory mechanism. The advent of unlimited gene purification by recombinant DNA cloning generalized Brown's approach. This gene-centric strategy unleashed a scientific revolution that unified genetics, developmental biology, and cell biology.

Born on 30 December 1931 in Cincinnati, Ohio, Brown attended Dartmouth College for 3 years and then earned his MD and an MS in biochemistry from the University of Chicago in 1956. After a yearlong internship, he spent 2 years as a National Institutes of Health fellow with neuroscientist Seymour Kety. He conducted postdoctoral studies with biochemist Jacques Monod at the Pasteur Institute in Paris and then moved to the Department of Embryology at the Carnegie Institution of Washington in Baltimore. He remained there for the rest of his career, becoming director of the department in 1976 and retiring with emeritus status in 2005.

When Brown was in medical school, embryo development was scarcely covered in the curriculum. He recognized that developmental biology was not yet a field but rather a set of questions. After the structure of the DNA double helix was published in 1953, Brown made it his scientific mission to merge DNA research, biochemistry, and embryology. He had the drive (as his opponents on the tennis court would attest), the self-confidence, and the courage to undertake this task.

Life at the Pasteur Institute inspired Brown, as work on bacterial operons (gene clusters regulated together) climaxed. He thrived in the lunchtime company of such luminaries as eventual Nobel laureates Monod, François Jacob, and André Lwoff. In contrast to the Pasteur Institute, Brown described the 1960 Carnegie Embryology Department as an "old-fashioned histology laboratory, or perhaps a museum." Yet, in its atmosphere of intellectual freedom and ample funding, he flourished.

The role of RNA as a direct product of genes had recently been recognized, and Brown decided to analyze RNA synthesis in frog embryos. He and John Gurdon found that a frog mutant lacking nucleoli produced neither of the major ribosomal RNAs (rRNAs), indicating that nucleoli were the normal site of rRNA production. Wondering why frog oocytes contained many more nucleoli than did frog somatic cells, Brown and Igor Dawid showed that the number of rDNA genes greatly increased ("amplified") during oogenesis to support the massive ribosome production needed for each oocyte, a discovery independently made by Joseph Gall.



The amplification and unusual base composition of rDNA made it possible to purify these genes according to their specific density, a feat first accomplished by Max Birnstiel in Geneva. Brown purified the genes encoding the smaller, more easily studied 5*S* rRNA genes and found that they could be accurately transcribed in vitro. Using recombinant DNA methods, Brown isolated plasmids containing the frog 5*S* genes, making them the first eukaryotic genes to be cloned. These methods led to Brown's astonishing discovery that, unlike bacterial operons, 5*S* gene transcription was controlled by sequences internal to the gene. Soon thereafter, he and Robert Roeder found that this control region binds a transcription factor, which they designated TFIIIA. The first gene-specific transcription factor described for eukaryotic cells, TFIIIA became the founding member of the zinc-finger class of protein domains.

Brown naturally grasped the motivations and idiosyncratic differences among young scientists, allowing him to find and me raw talent. As director, he transformed the Embryology Department into a place renowned for fostering faculty creativity and training future leaders. Brown hired distinctive individuals and gave them free rein to chart the department's course. His role was to support and encourage faculty so as to bolster the creativity of their research. "Life is too short not to focus on your best, most original ideas," he insisted. Those who worked there, including both of us, enhanced every aspect of their scientific capabilities. Generations of former students and postdoctoral fellows credit their long and successful careers to their time in Brown's department.

Brown made a department with only eight faculty intellectually larger by hiring young staff associates, mostly new PhDs, who were fully independent and adequately funded for 5 years. The success of Carnegie staff associates motivated multiple institutions to launch similar fellows programs. Countless first-time visitors remarked that the department was much smaller than they expected. A review committee once reported that it had detected no trace of administration in the department, a comment Brown—who took on the bulk of the administrative duties himself to prevent disruption to his staff—considered high praise. One of us (A.S.) succeeded Brown as director of the department in 1994 and continued to promote "the Carnegie style" that his department epitomized.

Brown favored interactions that were short, clear, and direct, whether in scientific discussion, written communication, or personnel evaluation. He would open meetings with a brief, riveting address that was insightful, humorous, and unconventional. His advice was sought by many institutions, including the Howard Hughes Medical Institute. His shrewd, unexpected suggestions were often highly influential.

Brown received multiple awards for his scientific achievements and leadership, including the 2012 Lasker-Koshland Special Achievement Award in Medical Science (shared with Tom Maniatis) for seminal contributions to understanding and manipulating genes and for mentoring young scientists. A champion of basic biological research who saw youth as its seed corn, Brown tirelessly empowered young scientists. He founded and for more than four decades nurtured perhaps his most fitting legacy: the Life Sciences Research Foundation, which has supported more than 700 postdoctoral researchers. Don Brown advanced the metamorphosis of developmental biology into a true discipline and made the scientific enterprise as a whole a more perfect servant of our highest ideals. ■

[1]University of Texas Southwestern Medical Center, Dallas, TX, USA. [2]Carnegie Institution for Science, Washington, DC, USA. Email: steven.mcknight@utsouthwestern.edu; spradling@carnegiescience.edu

10.1126/science.adj2815

POLICY FORUM

RESPONSIBLE RESEARCH

Researchers' responsibility to uphold Indigenous rights

Too often research brings harm to Indigenous peoples

By **Lawrence Ignace[1], Lauren Burton[1], Sara Mynott[1,2], Mairi Meehan[1,3], Erica Olson[1], Jade Steel[1], Jaime Ojeda[1,4], Sarah Harper[1], Luisa Ramirez[1], Dana Baker[1], Laurel Sleigh[1], Caitie Frenkel[1], Chris Rhodes[1], Natalie C. Ban[1]**

The United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), the most comprehensive international human rights instrument concerning Indigenous peoples, was adopted in 2007 by 144 countries to ensure protection of Indigenous rights and self-determination (*1*). While directed at states, UNDRIP should guide all levels of society. With governments lagging in its implementation, we argue that researchers—at universities, government institutions, consultancies, and elsewhere—have a responsibility to understand and advance these rights. As Indigenous and non-Indigenous researchers with backgrounds and experiences partnering with Indigenous peoples, we reviewed the articles of UNDRIP to identify those relevant to researchers. We summarized these into four themes: self-determination; free, prior, and informed consent; intellectual property; and engagement and learning. These articles provide a starting point for researchers to engage with and become allies in upholding Indigenous rights, in a way that supports Indigenous self-determination and sovereignty now and into the future.

Since 1948 the Universal Declaration of Human Rights (UDHR) has outlined fundamental rights to which all people are inherently entitled. Yet despite some benefits of the UDHR, such as influencing the development of international human rights law, its focus is on individuals, leaving many gaps for Indigenous peoples, who, despite UDHR, continue to be subjected to colonization, dispossession, discrimination, and genocide. Although there are jurisdictions that recognize Indigenous rights in law, such as in Bolivia and New Zealand, this is far from universal, and rights in law are often not adequately enforced. In response to all of this, leaders of Indigenous communities increasingly pressed for a distinct document that would go above and beyond UDHR and emphasize and enshrine unique rights to self-determination, lands, territories and resources, culture, identity, and language. The culmination of this massive effort was the UNDRIP.

Indigenous peoples have inherent rights "which derive from their political, economic, and social structures, and from their cultures, spiritual traditions, histories and philosophies, and especially their rights to their lands, territories and resources" (*1*). Although UNDRIP does not go so far as recognizing Indigenous sovereignty, it does summarize the minimum standard of rights for Indigenous peoples to be respected by states. UNDRIP is important because it attempts to right wrongs; equalize the uneven power relationship that resulted from colonialism, exclusion, and systemic racism; and navigate toward potential futures that uphold the rights and sovereignty of Indigenous peoples. UNDRIP is unique because it comprehensively articulates Indigenous rights relevant to these aims, unlike other agreements that affect, but are not explicitly focused upon, Indigenous peoples (e.g., Nagoya Protocol on Access and Benefit Sharing; Convention on Biological Diversity).

Implementation of UNDRIP by states varies. Some countries (e.g., Ecuador) have aligned their constitutions with UNDRIP, and others have applied it to respect Indigenous rights in court cases (e.g., Belize, Chile), whereas others are lacking any implementation. Researchers thus have a responsibility to act now to uphold UNDRIP because too often research continues to bring harm to Indigenous peoples and their territories. How research is conducted can have implications for Indigenous rights, self-determination, and sovereignty, especially when research takes place on Indigenous lands and waters and involves Indigenous peoples, their knowle wisdoms, cultural items, etc. (*2*). Some examples of violations of Indigenous rights by researchers include using biological samples for secondary research purposes without consent (*3*), and biopiracy of Indigenous knowledge for commercial exploitation (*4*). More subtle practices of scientific colonialism include non-Indigenous researchers implying ownership with phrases such as "my study sites" and "my Indigenous community," not including original (Indigenous) names when referring to species and places, or claiming to have made "discoveries" while ignoring Indigenous knowledge. Such practices are common, amplifying the need for researchers to ensure that Indigenous peoples have the foresight, power, and authority to determine what, how, and why research happens in their territories.

UNDRIP implementation by researchers exists within three interrelated spheres of control, influence, and interest (*5*). Researchers hold power, privilege, and responsibility as they have control over where, when, how, and why their research is carried out, throughout all stages in the research process, at the project and program levels (*5*). Whereas these aspects of research are within the direct control of researchers, others can be influenced, but not controlled, such as how research is conducted at the institutional level (e.g., changing knowledge, attitudes, and skills through interpersonal relationship building; contributing to or chairing committees; and supporting institutional governance reform). Beyond their spheres of control and influence lies their sphere of interest, where decisions within the sphere of control (e.g., sharing research findings with practitioners and policy-makers) interact with governance and policy decision-making and can affect concrete social, economic, and environmental changes. How much researchers can affect the spheres of control, influence, and interest depends on the individual and the autonomy of their position (e.g., a research assistant may have little control, whereas a university president can shape policy).

RESEARCH-RELEVANT THEMES

The researcher-relevant themes that we identified in UNDRIP are especially important to Indigenous peoples because of the long-standing attempts by colonizers and settlers to erase their inherent rights and delegitimize their worldviews and relationality.

Self-determination

A core theme in UNDRIP is that "Indigenous peoples have the right to self-determination" (Article 3) and includes Indigenous peoples' right to self-government (Article 4), main-

[1]School of Environmental Studies, University of Victoria, Victoria, BC, Canada. [2]Centre for Marine Socioecology, Hobart, Tasmania, Australia. [3]Memorial University of Newfoundland, St. Johns, NL, Canada. [4]Cape Horn International Center for Global Change Studies and Biocultural Conservation (CHIC), Universidad de Magallanes, Chile. Email: lawrenceignace@uvic.ca

taining and strengthening their institutions (Article 5), the right to practice and revitalize cultural traditions and customs (Article 11.1), and a right to participate in decision-making (Article 18). This means that upholding UNDRIP requires conducting research in a way that allows Indigenous peoples to determine what and how research happens in their territories. Indigenous peoples draw upon origins from their worldviews, lands, and creation stories, which form their moral and societal responsibilities to the environment and their communities (*6*, *7*). The implication for research in support of Indigenous self-determination will be the acceptance and manifestation of Indigenous ways of knowing and being across all research stages and disciplines. That researchers must recognize and engage with Indigenous ways of knowing does not necessarily mean it must then replace other ways of knowing. For example, the concept of Etuaptmumk (Mi'kmaw for "Two-Eyed Seeing") provides a conceptual framework for embracing multiple perspectives within a system (*8*).

Researchers should be guided by UNDRIP in creating space for equitable relationships and places of sharing and learning. Researchers can create opportunities for Indigenous peoples to freely determine their own research agendas by co-creating research objectives, questions, and approaches. In relationship with their Indigenous community partners, researchers can bring to the forefront Indigenous epistemologies and Indigenous-led approaches that support Indigenous self-determination. For Indigenous peoples, these steps can help with the continued survival of their cultures, languages, and lands and ideally can support Indigenous sovereignty (*9*, *10*).

Free, prior, and informed consent

UNDRIP states that Indigenous people must give "free and informed consent prior to the approval of any project affecting their lands or territories and other resources" (Article 32.2), and "before adopting and implementing [...] measures that may affect them" (Article 19). Researchers must ensure that free, prior, and informed consent is obtained for all stages of the research process, as is necessary for Indigenous self-determination. This goes beyond informed consent of individual research participants—which is required for all human subjects—to include ongoing consent to conduct research in Indigenous territories. Ensuring consent requires that researchers accept that Indigenous peoples may not be interested in the proposed research, i.e., researchers must be willing to accept "no" as an answer. Additionally, priorities may shift over the course of a project; informed consent must be ongoing and revisited. If consent is withdrawn during the course of a research project, and the research cannot be reshaped to continue collaboration, researchers need to communicate with relevant parties (e.g., funders, other collaborators) to identify next steps. Ensuring consent, and preventing harms, require careful consideration of how data collection, new knowledge developed, and outputs generated could infringe on Indigenous peoples' rights. It also requires understanding the Indigenous context, including historical experiences and desired futures. Repositioning Indigenous priorities within research can build beneficial and lasting relationships, address Indigenous needs, strengthen reconciliation, and improve research.

tural heritage, traditional knowledge, and traditional cultural expressions" (Article 31.1). Also emphasized is redress for "cultural, intellectual, religious and spiritual property taken without their free, prior and informed consent or in violation of their laws, traditions, and customs" (Article 11.2).

Indigenous data sovereignty conveys the right of an Indigenous group to reclaim and govern the use of its own intellectual property, including data collection, organization, and storage (*11*). The importance of data sovereignty is further affirmed by recent development of Indigenous data governance frameworks [e.g., (*11*)] and tools, which include procedures that researchers must adopt to increase transparency, protect Indigenous intellectual property, and align with specific

Luis Levil, a Huilliche-Chilote fisher, and PhD student Jaime Ojeda feed hake offal to seabirds in southern Patagonia, Chile, as part of a research project about the hake fishery.

Engaging with the highest level of ethical standards, and analyzing risk, are important to identify potential research impacts at the outset. Many Indigenous communities have created their own ethics guidance to create equitable research processes [e.g., (*10*)], including how to obtain their free, prior, and informed consent. Indigenous peoples' guidance should be followed where research is undertaken in their respective territories.

Intellectual property

Ensuring Indigenous peoples control their own intellectual property is reflected in several Articles of UNDRIP. For example: "Indigenous peoples have the right to maintain, control, protect and develop their cultural

Indigenous protocols (e.g., codeveloping agreements for research collaboration and data sharing). Researchers have a responsibility to protect Indigenous data because governing their own intellectual property is fundamental to self-determination (*11*).

Engagement and learning

UNDRIP states that Indigenous peoples have a right to establish educational systems "in a manner appropriate to their cultural methods of teaching and learning" (Article 14.1); to access "all levels and forms of education of the State without discrimination" (Article 14.2); to transmit knowledge outside of educational institutions and to "retain their own names for communities, places and persons"

(Article 13.1); to be "appropriately reflected in education and public information" (Article 15.1); and to "understand and be understood" in policymaking (Article 13.2) and in media (Articles 16.1 and 16.2).

Researchers in all disciplines must respect Indigenous methodologies and recognize the knowledge that already exists in areas where they study. Researchers have a responsibility to share findings with participants so they can be transmitted to future generations and should do so in an understandable format and without institutional barriers. Indigenous knowledge and perspectives should be appropriately represented in the work researchers share with their peers or the public. Indigenous researchers bear considerable responsibilities of representing Indigenous voices in research. Non-Indigenous researchers can support Indigenous peoples by improving capacity for collaboration and innovation to support self-determination.

Good practice in all community-engaged research—which is essential when working in Indigenous territories—is to share project updates and results frequently with the Indigenous communities with whom they collaborate; publish results in open-source journals and nonacademic formats whenever possible and where permission is granted; acknowledge relevant Indigenous community, history, knowledge, and place names; and support Indigenous colleagues and emerging scholars through supporting grant writing, hiring Indigenous research assistants, and creating mentorship opportunities.

MOVING FORWARD

Researchers have a responsibility to uphold Indigenous rights by reflecting on the articles of UNDRIP as they relate to their research. Researchers are well-positioned to lead by example not only in their own work, but also in influencing the related institutions and funders that support research. Working with and for Indigenous peoples in a good way requires understanding the effects of colonial histories on Indigenous peoples and the contribution of research and research institutions to past and ongoing colonial legacies, as well as the systemic barriers that continue to exist that disadvantage Indigenous peoples, recognizing existing capacities and strengths, and advocating for change.

Researchers control and shape how they carry out their own work and hence there is no excuse for not upholding Indigenous rights. To uphold UNDRIP, researchers must adopt a transdisciplinary research approach that engages Indigenous methodologies, where appropriate, instead of prioritizing Western research framing (*12*). Indigenous methodologies are based on the worldview, values, and epistemology of Indigenous peoples, and hence are diverse. Often, they challenge dominant Western methodologies that marginalize, misrepresent, or silence Indigenous voices and perspectives. The potential misalignment between the two arises from fundamental differences in epistemologies, ontologies, and research methodologies. Frictions might occur as a result of siloed mindsets and Western methodologies that prioritize supposed objectivity and quantitative data over Indigenous ways of knowing, or when Indigenous frameworks challenge Western notions of ownership, consent, or data sovereignty. By rebalancing research dynamics and shifting power structures, knowledge coproduction becomes the standard, creating and maintaining ethical space (*9*). It takes time to conduct research respectfully through meaningful dialogue and relationships. Much guidance exists on partnering with Indigenous peoples (*7*–*13*), and researchers need to do their own learning and reflecting prior to engagement.

Researchers can play an important role in influencing institutions and education spaces (e.g., universities, government research institutes, funding agencies) to embed Indigenous self-determination in research practices (*6*) by compelling institutions to develop processes that ensure UNDRIP is followed. For example, much like requiring human and animal ethics applications, and permission to access private or state-owned lands, institutions must mandate researchers to follow protocols, conditions, and permissions set and required by the Indigenous peoples on whose territory (whether State-recognized or not) they wish to conduct research, and to only proceed with research when permissions are granted. Institutions must ensure that processes and resources are available, especially for Indigenous peoples to build and maintain the capacity to respond to research requests and participate as equal or leading research partners. Furthermore, researchers have the power to influence funders in adapting funding mechanisms that enable Indigenous peoples to lead research in their territories. Leadership is especially needed from senior researchers and disciplinary leaders to transform the performance measures and cultures dominant in many research institutions that hinder the development of research practices consistent with UNDRIP. This should not be a burden imposed solely on young researchers and Indigenous scholars and communities.

Implementing UNDRIP through researchers' spheres of control and influence will set the stage for supporting broader social, economic, and environmental desirable changes within their sphere of interest. Researchers can affect systemic change by supporting Indigenous sovereignty—for example, through sustained research partnerships and following Indigenous-led approaches. As such, researchers can contribute to broader policy changes needed to ensure that the inherent rights of Indigenous peoples are recognized in all aspects of society. Examples range from framing fisheries management with Indigenous lenses (*13*) to the Ecuadorian and Bolivian constitutions being based on Indigenous epistemologies (*14*).

It takes long-term commitment and creativity to move beyond the tried and tired research practices common among today's colonial research paradigms (*12*, *15*). UNDRIP is a good starting point, but we also recognize that serious concessions continue to be required by Indigenous peoples within UNDRIP. For instance, although UNDRIP recognizes Indigenous self-determination, it does not go so far as to support Indigenous sovereignty. Indigenous ontological and epistemological frameworks offer different approaches in reconsidering more innovative research learning opportunities to guide our shared future. The research community has an opportunity and responsibility to work more closely with Indigenous peoples to create a more equitable and sustainable future based on the rights and concepts within UNDRIP, and beyond. ■

REFERENCES AND NOTES

1. United Nations "The United Nations Declaration on the Rights of Indigenous Peoples" (2018).
2. C. Wong, K. Ballegooyen, L. Ignace, M. J. G. Johnson, H. Swanson, *Facets* **5**, 769 (2020).
3. R. Dalton, *Nature* **420**, 111 (2002).
4. C. Cottrell, *Environ. Sci. Policy* **135**, 162 (2022).
5. B. Belcher, J. Halliwell, *Humanit. Soc. Sci. Commun.* **8**, 183 (2021).
6. W. Ermine, R. Sinclair, B. Jeffery, "The Ethics of Research Involving Indigenous Peoples." Report of the Indigenous Peoples' Health Research Centre to the Interagency Advisory Panel on Research Ethics (Saskatoon, 2004).
7. D. McGregor, *Am. Indian Q.* **28**, 385 (2004).
8. A. J. Reid *et al.*, *Fish Fish.* **22**, 243 (2021).
9. N. Latulippe, N. Klenk, *Curr. Opin. Environ. Sustain.* **42**, 7 (2020).
10. Kitasoo/Xai'xais Stewardship Authority, "Informing First Nations Stewardship with Applied Research: key questions to inform an equitably beneficial and engaged research process" (Kitasoo/Xai'xais First Nation, 2021).
11. C. Callison, A. Ludbrook, V. Owen, K. Nayyer, *KULA: Knowledge Creation, Dissemination, and Preservation Studies* **5**, 1 (2021).
12. L. T. Smith, *Decolonizing Methodologies: Research and Indigenous Peoples* (Zed Books, ed. 3, 2021).
13. J. K. Riddell, A. Salamanca, D. J. Pepler, S. Cardinal, O. McIvor, *Int. Indig. Policy J.* **8**, (2017).
14. P. Lupien, *Democratization* **18**, 774 (2011).
15. J. T. Johnson *et al.*, *Sustain. Sci.* **11**, 1 (2016).

ACKNOWLEDGMENTS

The authors thank the Quadra Centre for Coastal Dialogue for hosting a meeting at which ideas contained in this paper were developed, and thank N. Talloni for input. Funding was provided by Natural Sciences and Engineering Research Council of Canada E.W.R. Steacie Fellowship 556127-2021 (N.C.B.) and Natural Sciences and Engineering Research Council of Canada Discovery grant 4249-2020 (N.C.B.). All authors contributed to all stages of the paper development, with L.I. leading on the conceptualization, N.C.B. leading the original drafting, and N.C.B. providing funding.

10.1126/science.adh4470

A client waits for the Affordable Care Act website to come back online on 31 March 2014.

TECHNOLOGY POLICY

Delivering democracy

Better digital tools can improve government services and strengthen nations

By **Camille Crittenden**

In *Recoding America*, former US deputy chief technology officer for government innovation Jennifer Pahlka offers a remarkably compelling, inspiring, and entertaining read about government technology. Her behind-the-scenes look at entrenched US systems of government service delivery illustrates how she and her teams have overcome bureaucratic sludge to improve lives. As the founder and former executive director of Code for America, a nonprofit organization that aims to strengthen democracy by advancing digital tools for government service, she is well qualified to address these topics.

Recoding America is as much about clashing cultures as it is about software development methodology and technology platforms. Pahlka describes the historical hierarchy between policymakers ("big picture thinkers") and IT teams ("mechanicals") and offers evidence of how closer collaboration between these groups would help lawmakers achieve the policy outcomes they seek.

For instance, when California's Proposition 64 came into effect in 2016, it allowed those convicted of certain marijuana-related felonies to petition for their records to be sealed. Sweeping in its intent, it left the burden of petitioning on those who had been convicted. Pahlka describes how she and her team at Code for America worked with the San Francisco district attorney to develop tools to automate record expungement, resulting in more than 8000 records being cleared in San Francisco alone.

Automating the implementation of other, similar policies has proven more difficult. California's Proposition 47, for example, reduced felony convictions to misdemeanors for those convicted of commercial burglary in amounts of $950 or less. However, there is no specific legal code that captures commercial burglaries or the estimated value of stolen property. Finding redress for those convicted is a laborious, time-consuming process that is resistant to automation. The result is that thousands of low-level offenders will have difficulty finding housing or employment because of convictions that should have been erased.

Recoding America: Why Government Is Failing in the Digital Age and How We Can Do Better
Jennifer Pahlka
Metropolitan Books, 2023. 336 pp.

Government IT teams must often work with outdated legacy hardware, software, and data storage systems. Implementation timelines for tech upgrades lag those in the private sector, and the process can take 12 to 18 months or longer. Technology capabilities can change drastically over this period, meaning that even before a technical development is undertaken, it may be eclipsed.

Pahlka advocates that government entities adopt the more current "agile" approach, which involves multiple rounds of testing and iteration before a final product is delivered, and she shows how a lack of tech agility can cause real-world hardship. A harrowing example of such a shortfall can be found in the stories of families separated at the border under the Trump administration. The computer registration system used by US Border Patrol officials provided no way to associate ID numbers assigned to unaccompanied minors with those of adult family members. To update the database architecture and software would have required a new competitive procurement process, which could take 3 to 5 years. Officials resorted to placing sticky notes with identifying information on infants. It would be months before many of the children were reunited with their parents.

Throughout the book, Pahlka maintains a consistent clarity of purpose, focusing on how updating a given digital process can improve service delivery and thus strengthen democracy. Reducing administrative burden (liked by the Left) and reducing regulatory burden (favored by the Right) could, she believes, unite lawmakers across the political spectrum.

Pahlka shares hard-won lessons for how to streamline workflows and simplify processes for end users and agencies alike. Gather only the information necessary for a given purpose, she advises. When applying for food assistance, for example, individuals need not be asked whether they own a burial plot (as they currently are in many states, as an indicator of the applicant's assets). And design for mobile first. Cell phones are the most frequent interface used by low-income residents. "Paperwork favors the powerful," she writes, reminding readers that those eligible for government benefits may not have the time and ability to complete arduous documentation. Reducing paperwork will thus enable families with fewer resources to attain the assistance they need.

Pahlka's view of government operations is expansive, and I would have been interested to learn more about her perspective on other timely topics, including the notable presence of female leaders in civic technology and the opportunities and challenges presented by generative artificial intelligence for government services. I highly recommend this valuable and enlightening book to anyone in public-sector IT, as well as to policy-makers and their staff. ■

10.1126/science.adi5581

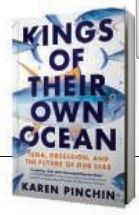
The reviewer is at the Center for Information Technology Research in the Interest of Society (CITRIS) and the Banatao Institute, University of California, Berkeley, CA 94720, USA. Email: ccrittenden@berkeley.edu

MARINE ECOLOGY

The ocean's disappearing giants

A journalist documents the struggle to catch, tag, and restore populations of Atlantic bluefin tuna

Kings of Their Own Ocean: Tuna, Obsession, and the Future of Our Seas
Karen Pinchin
Dutton, 2023. 320 pp.



By **Olaf P. Jensen**

We observe all animals through the lens of our human perspective, but some are especially obscured by the clouded glass of symbolism and story. So it is for the bluefin tuna, a species overfished to the degree that it has become a living shorthand for man's wastefulness and greed. Few of us have seen this fish swimming free in its natural environment; they are rare even in aquaria. Those who see it at all see its death—the staccato of its crescent moon tailfin beating a ship's deck, power without purchase.

In *Kings of Their Own Ocean*, Karen Pinchin tells the stories of people whose lives intersect with Atlantic bluefin tuna, one of three bluefin tuna species found in the world's oceans. Most prominently, there is Al Anderson, a salty charter captain from Rhode Island. Al preaches the gospel of catch-tag-and-release and makes converts of his fishing clients, many of whom come aboard expecting to bring home sashimi but instead return to the dock with a tagging card full of data about the fish they released. There are also fascinating appearances by the Reverend Sun Myung Moon and his followers, who took over bluefin tuna fishing and buying in Gloucester, Massachusetts, in the 1970s; Spanish trap fishermen who scuba dive with penned tuna to herd them deeper into the traps; and tuna researcher Molly Lutcavage, a pioneer in the use of electronic tags to track movements of large pelagic fish.

The thread that ties these stories together is Amelia, a one-in-a-million fish that was tagged and recaptured multiple times during her 15- or 16-year life as she traveled from the US Northeast to a Spanish fish trap in the Mediterranean, where she met her end. Pinchin conveys the excitement of recapturing a tagged fish and the inevitable wonder about its life since release.

I still, after having received hundreds of returned fish tags, rush to my laptop to look up tag numbers and find out when and where a fish was tagged, how much it grew, and, the unanswerable, what did it experience while it was "at large"? In Amelia's case, that experience is known in more geographic detail, as she carried an electronic tag that tracked her movements for several months.

The people in *Kings of Their Own Ocean* are portrayed in all their rich human complexity, a quality that sets this book apart from many of the other nonfiction works focused on a single fish species that followed Mark Kurlansky's 1997 international bestseller, *Cod: A Biography of the Fish That Changed the World*. However, Pinchin presents the overfishing of bluefin tuna mostly as a simple tale of greed. The "bad guys" (Japanese fishery management delegations and commercial fishers) and the "good guys" (anglers who practice catch-and-release and conservationists raising the alarm) might as well wear black and white hats. Readers are introduced to a couple more obscure villains, including the concept of maximum sustainable yield and the partitioning of Atlantic bluefin tuna into two geographic stocks, both of which Pinchin blames for driving overfishing, but this part of the book is mostly well-trodden ground. If there is something new here, it is the meta-story of how conservationist and author Carl Safina managed to equate bluefin tuna with human greed in the public mind and thus launch a consumer-driven marine conservation movement.

But fisheries are more complex than this—which is what makes them so interesting to study. The sustainability of tuna fisheries is driven not only by the decisions of the fishermen but also by biology, economics, and the institutions that attempt to manage the whole social-ecological system.

Tuna fishing is expensive, and the biological and economic equilibrium point is a function of costs, revenue, and the ability of the species to withstand harvest (*1*). Through well-meaning but misguided fuel subsidies for fishermen, many national governments have driven this "bionomic equilibrium" toward more fishing and fewer tuna left in the sea. The upshot is that across species and oceans, the conservation status of tuna can be reasonably well predicted from two variables: their generation time and market price (*2*). The valuable and relatively long-lived Atlantic bluefin tuna suffers on both counts.

Yet all is not lost. The most recent assessment of the Western Atlantic stock (*3*) shows the population recovering in a long, slow arc starting around 2007 after fishing mortality rates were finally brought down to consistently lower levels. While it is too soon to pop the champagne, Pinchin ends on a note of well-founded optimism that bluefin tuna management might finally be turning a corner on the era of unrestrained overfishing. ■

Bluefin tuna are hauled into a boat near the Barbate coast in Spain.

PHOTO: PABLO BLAZQUEZ DOMINGUEZ/GETTY IMAGES

The reviewer is at the Center for Limnology, University of Wisconsin–Madison, Madison, WI 53706, USA.
Email: ojensen@wisc.edu

REFERENCES AND NOTES

1. H. S. Gordon, *J. Polit. Econ.* **62**, 124 (1954).
2. B. B. Collette *et al.*, *Science* **333**, 291 (2011).
3. International Commission for the Conservation of Atlantic Tunas, Report of the 2021 Western Bluefin Stock Assessment Meeting (2021); https://apps-st.fisheries.noaa.gov/sis/docServlet?fileAction=download&fileId=7653.

10.1126/science.adi5203

LETTERS

Edited by **Jennifer Sills**

Undue publicity for flawed fraud detector

In the News In Depth story "Fake scientific papers are alarmingly common" (9 May, https://scim.ag/2KD), J. Brainard describes a medRxiv preprint by Sabel *et al.* (*1*) alleging that nearly 30% of all scientific articles may be the fake products of paper mills. Fraudulent papers are indeed a concern, but this story sensationalizes the study's findings while downplaying its substantial flaws. In doing so, Brainard unjustly tarnishes the scientific enterprise and causes undeserved harm to the reputation of scientists from the countries singled out in the preprint.

Sabel *et al.* propose a classifier for identifying fake papers. Rather than training this classifier using machine learning, the researchers preselected just three features to flag suspected fakes: whether the authors use private email addresses, whether they are affiliated with a hospital, and—unmentioned by the News story—whether the team lacks international collaborators. Brainard eventually cites details from the preprint showing that the algorithm misclassified more than a third of real papers as fake. This unacceptable false-positive rate indicates that the method has failed.

The structure of the Sabel *et al.* study demonstrates how such classifiers can be used to reinforce discrimination and inequity. Because papers without international coauthors, as well as authors without institutional email addresses, are deemed untrustworthy, the classifier is more likely to identify papers from Asia and the Global South, where international collaborations are less common (*2*) and personal email addresses are more frequently used (*3*). Yet Sabel and colleagues use the output of this classifier to argue that fraud is more common among authors from these countries, without controlling for these well-documented international differences. Such circular reasoning unjustly disparages authors from these countries.

The algorithm that *Science* has heralded is fundamentally biased in its implementation, with potentially racist consequences. Using a highly biased algorithm even as a preliminary screen is unacceptable without carefully considering issues of fairness and algorithmic harm that arise from its use. No such consideration is given, and the News story gives only passing mention to the risks. *Science* should be on the vanguard of drawing attention to how machine intelligence can encode racism while perpetuating and exacerbating traditional inequities, not an accessory to such wrongs.

Carl T. Bergstrom[1]* and C. Brandon Ogbunugafor[2]
[1]Department of Biology, University of Washington, Seattle, WA, USA. [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA.
*Corresponding author.
Email: cbergst@u.washington.edu

REFERENCES AND NOTES

1. B. A. Sabel, E. Knaack, G. Gigerenzer, M. Bilc, medRxiv 10.1101/2023.05.06.23289563 (2023).
2. A. Gazni, C. R. Sugimoto, F. Didegah, *J. Am. Societ. Info. Sci. Technol.* **63**, 323 (2012).
3. S. Shen, R. Rousseau, D. Wang, *Scientometrics* **115**, 1039 (2018).

10.1126/science.adi7104

Advancing equity and integrity in research

In the News In Depth story "Fake scientific papers are alarmingly common" (9 May, https://scim.ag/2KD), J. Brainard draws attention to a number of new tools and technologies that researchers and publishers are developing in an effort to increase the integrity of research. These tools primarily work by automating the identification of fake authors and fake papers with the use of a set of variables common in publications, such as author email address and affiliation. The goals of these technologies are valid and important, but the scientific community needs to act cautiously when determining which variables are used to identify and flag individuals and manuscripts.

A private email address could be considered a flag for further scrutinizing a publication, but private email addresses are often used by legitimate researchers, many of whom are nonnative English speakers or use non-Latin writing scripts. In countries such as India and China, more researchers use personal email addresses than institutional ones (*1*). Judging the legitimacy of research based on this flag coupled with a hospital affiliation, as one medRxiv preprint Brainard cites does (*2*), without any follow-up practice that accounts for language or cultural differences, would be a disservice to scientific advancement globally.

More rigorous author and manuscript evaluation needs to verify individuals and institutions by leveraging not

ILLUSTRATION: SARA GIRONI CARNEVALE

just email domain and affiliations but also author disambiguation tools and persistent identifier registries, such as Open Researcher and Contributor Identifiers (ORCIDs) and the Research Organization Registry. An additional analysis of publishing history—including the number of publications, coauthorship, and a network analysis—should be applied to create signals for additional checks on the author of a manuscript. These results should then be balanced with knowledge and sensitivity. Graduate students and new faculty with a light publication history are not fake authors, just new to the field.

Ensuring research integrity requires reducing inequities and gatekeeping in the publishing ecosystem. Some of the tools cited by Brainard use limited checks that prematurely categorize and ultimately miscategorize researchers as nefarious actors. Both the proposed tool and the choice by *Science* to promote it contribute to furthering inequities rather than improving trust in science.

The Committee on Publication Ethics (COPE) has developed sensible guidance (*3*) for editors and publishers on how to navigate any potential paper-mill case once it has been identified. Better evaluation mechanisms will allow the scientific community to simultaneously advance equity and integrity goals.

Leslie McIntosh[1]* and Cynthia Hudson Vitale[2]
[1]Digital Science, London, UK. [2]Association of Research Libraries, Washington, DC, USA.
*Corresponding author.
Email: l.mcintosh@digital-science.com

REFERENCES AND NOTES

1. S. Shen, R. Rousseau, D. Wang, *Scientometrics* **115**, 1039 (2018).
2. B. A. Sabel, E. Knaack, G. Gigerenzer, M. Bilc, medRxiv 10.1101/2023.05.06.23289563 (2023).
3. COPE, "Addressing concerns about systematic manipulation of the publication process" (2023).

10.1126/science.adi7867

Editor's note

We thank Bergstrom and Ogbunugafor, as well as McIntosh and Hudson Vitale, for pointing out the limitations of tools for detecting fake papers such as the one developed by Sabel *et al.* Our story states prominently, in the fourth paragraph, that Sabel *et al.*'s method "isn't a perfect solution, because of a high false-positive rate." Later we specify the high false-positive numbers. Far from heralding or sensationalizing the tool, we presented it as a rough indicator of a real problem.

Tim Appenzeller
News Editor, *Science*

10.1126/science.adj3681

Where Science Gets Social.

AAAS.ORG/COMMUNITY

AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

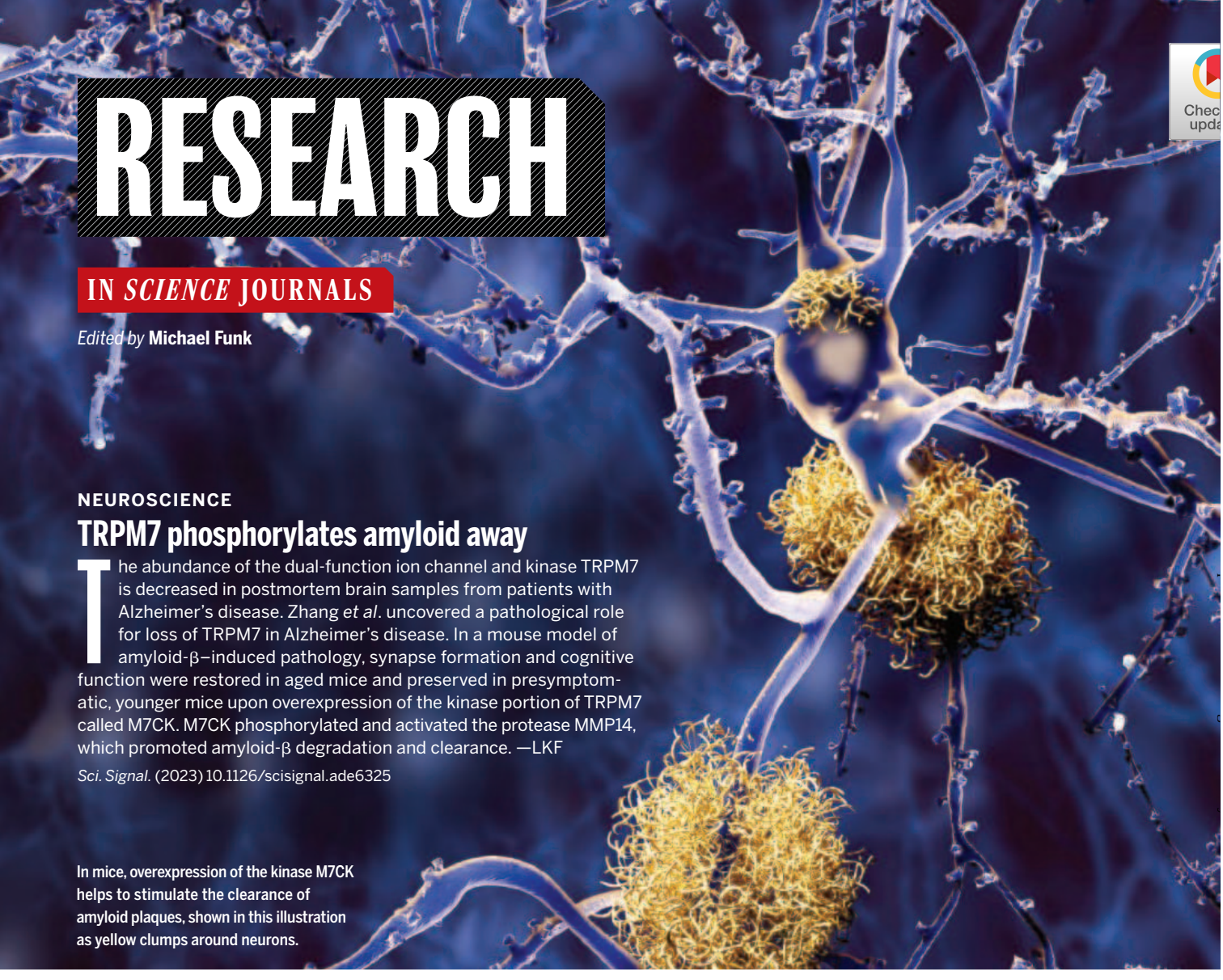
Member COMMUNITY AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



In mice, overexpression of the kinase M7CK helps to stimulate the clearance of amyloid plaques, shown in this illustration as yellow clumps around neurons.

NEUROSCIENCE

TRPM7 phosphorylates amyloid away

The abundance of the dual-function ion channel and kinase TRPM7 is decreased in postmortem brain samples from patients with Alzheimer's disease. Zhang *et al*. uncovered a pathological role for loss of TRPM7 in Alzheimer's disease. In a mouse model of amyloid-β–induced pathology, synapse formation and cognitive function were restored in aged mice and preserved in presymptomatic, younger mice upon overexpression of the kinase portion of TRPM7 called M7CK. M7CK phosphorylated and activated the protease MMP14, which promoted amyloid-β degradation and clearance. —LKF

Sci. Signal. (2023) 10.1126/scisignal.ade6325

QUANTUM OPTICS

Controlling quantum randomness

The state of quantum mechanical systems fluctuates. In optics, these fluctuations can be harnessed for the generation of truly random bit strings. The random fluctuations can also be used in probabilistic computing, but this requires the probability distributions to be controllable. Using the two-state phase output from an optical parametric amplifier as a "bit," Roques-Carmes *et al*. showed that the bit probabilities changed in response to a small bias field injected into the system. By changing the attenuation level of the bias field, the authors were able to traverse the continuous space between perfectly random and deterministic state selection. The approach provides the prospect of using vacuum fluctuations as a source of controllable randomness for photonic probabilistic computing. —ISO

Science, adh4920, this issue p. 205

SOLAR CELLS

Less penetrating but more stable

The thermal stability of three-dimensional perovskite solar cells can be improved by adding ammonium ligands that create a two-dimensional perovskite capping layers, but such ligands are prone to intercalation into the bulk. Park *et al*. showed that the smallest aromatic ligand, anilinium, had the lowest ligand reactivity with three-dimensional perovskites because of steric hindrance near the ammonium group, and a fluorinated derivative of this ligand created a robust interface structure. Encapsulated solar cells maintained 85% of their power conversion efficiency of about 20% at 85°C and 50% relative humidity after about 1600 hours of maximum power point operation. —PDS

Science, adi4107, this issue p. 209

MOLECULAR BIOLOGY

Corepressors and enhancers form partners

The regulation of gene expression involves different signals to switch transcription on or off. Gene activation has received widespread attention, but fewer studies have looked at how repressive transcription factors and their partners known as corepressors silence transcription. Jacobs *et al*. now show that not all corepressors can silence all gene-activating regions, called enhancers. Instead, corepressors work selectively, depending on specific enhancer characteristics such as chromatin and sequence patterns, and altering these motifs can make enhancers more or less sensitive to specific corepressors. This finding uncovers another layer of gene expression control and highlights the interplay between distinct mechanisms of transcriptional activation and repression. —DJ

Science, adf6149, this issue p. 198

CHATGPT

Generative AI as a writing aid

Automation has historically displaced human workers in factories (e.g., automotive manufacturing) or in performing routine computational tasks. Will generative artificial intelligence (AI) tools such as ChatGPT disrupt the labor market by making educated professionals obsolete, or will these tools complement their skills and enhance productivity? Noy and Zhang examined this issue in an experiment that recruited college-educated professionals to complete incentivized writing tasks. Participants assigned to use ChatGPT were more productive, efficient, and enjoyed the tasks more. Participants with weaker skills benefited the most from ChatGPT, which carries policy implications for efforts to reduce productivity inequality through AI. —EEU

Science, adh2586, this issue p. 181

TUMOR IMMUNOLOGY

Instructing tumor cell turncoats

Cancer cells down-regulate antigen presentation to evade recognition by the immune system. Zimmermannova *et al.* reprogrammed cancer cells into tumor-derived antigen-presenting cells (tumor-APCs). Mouse or primary human cancer cells transduced with a set of transcription factors acquired a stable dendritic cell phenotype, could present endogenous tumor antigens to stimulate the effector function of $CD8^+$ T cells, and showed decreased tumorigenicity. In mouse models, intratumoral injection of tumor-APCs decreased tumor growth, increased the survival of tumor-bearing mice, and enhanced responses to immune checkpoint inhibitors. This study highlights that future immunotherapies aimed at reprogramming cancer cells in situ could combine reversing malignancy with mobilizing the cellular antigen-presenting machinery to enhance the activity of $CD8^+$ T cells toward tumors. —SHR

Sci. Immunol. (2023) 10.1126/sciimmunol.add4817

QUANTUM NETWORKS

Building a multichip quantum network

The realization of a scalable quantum entanglement network relies on the development of a scalable architecture and on incorporating hardware components that are mass manufacturable. Zheng *et al*. proposed and demonstrated a scalable architecture using photonic hybrid multiplexing. They realized a toolbox of integrated nanophotonic hybrid multiplexing devices that can be manufactured using standard industrial fabrication processes. Their chip-based quantum entanglement network could find important applications in quantum communications, metrology, and distributed quantum computing. —ISO

Science, adg9210, this issue p. 221

SUPERCONDUCTIVITY

What sets the transition temperature?

The mechanism of high-temperature superconductivity in copper oxide materials remains a mystery more than 30 years after its initial discovery. One way to shed light on this is to look for correlations between different observables in cuprate families. Wang *et al*. used electron energy-loss spectrometry to measure the charge transfer gap in a class of bismuth-based cuprates with varying numbers of copper oxide planes in the unit cell. The measured gap was negatively correlated with the maximum transition temperature, providing a design principle for further increasing the working temperature of cuprate superconductors. —JS

Science, add3672, this issue p. 227

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

Spotted-winged fruit fly in flight over a strawberry, one of the many fruit species that this insect's larvae feeds on

EVOLUTION

A hearty eater

Many plant-eating insects are picky eaters and specialize on a single host plant. Some species do have generalist appetites, but whether they use the same biochemical processes as specialists to convert different foods into energy and other bioactive molecules is not well understood. Olazcuaga *et al.* fed a generalist species, the spotted wing fruit fly, on blackcurrant, cherry, cranberry, and strawberry separately and then measured the chemical compositions of the fruits and the flies. The authors found that the same metabolism was used by flies to convert fruit into energy regardless of which species they ate. Consequently, some harmful chemicals unique to a fruit were not detoxified and remained in the body. There may thus be a negative tradeoff for the advantages of being a generalist. —DJ *eLife* (2023) 10.7554/eLife.84370

NEUROSCIENCE

Unique development of the insular cortex

During fetal development, the human brain expands to form cortical lobes that fold into gyri (ridges) and sulci (grooves). Models of brain folding can explain the genesis of individual gyri and sulci. However, these models exclude the formation of the insula, a region deep inside the cortex involved in

PHOTO: NATURE PICTURE LIBRARY/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

PLANT SCIENCE

Metabolic enzymes moonlighting on histones

The methylation state of histones affects chromatin structure and gene expression. Histone demethylase activity influences the distribution of these regulatory methylation marks. Huang *et al*. found that an enzyme called α-ketoglutarate dehydrogenase (KGDH) binds to and inhibits the activity of histone demethylases (see the Perspective by Taylor). KGDH is normally associated with the respiratory tricarboxylic acid cycle in mitochondria, but its substrate, α-ketoglutarate, is also required for histone demethylase activity. The KGDH enzyme enters the nucleus of *Arabidopsis* cells and can interact directly with both histone demethylases and chromatin. As a result, both histone demethylation and gene expression are altered by KGDH, providing an alternative role for KGDH beyond respiratory metabolism. —MRS

Science, adf8822, this issue p. 179; see also adi7577, p. 125

ANTIMICROBIAL PROTEINS

Lysis by phage protein E

With the progressive loss of effective antibiotics against many bacterial pathogens, phage therapy has attracted renewed interest as a means of treating infections. Some phages generate antimicrobial proteins that result in lysis of their host cells. Orta *et al*. determined structures of the "YES" complex, consisting of the phage protein E and two proteins involved in cell wall biosynthesis, MraY and SlyD (see the Perspective by Balbach and Stubbs). Protein E blocks a substrate-binding site in MraY, thus blocking its essential enzyme activity. A better understanding of this mechanism may be useful in developing effective phage therapeutics. —MAF

Science, adg9091, this issue p. 180; see also adi7571, p. 126

TOPOLOGICAL MATTER

Capturing the effects of the quantum metric

The quantum metric, which quantifies the distance between neighboring wavefunctions in a quantum system, has been predicted to affect many physical responses. One example is the predicted nonlinear anomalous Hall effect in antiferromagnets with particular symmetries. Gao *et al*. observed this effect in a structure consisting of an even number of layers of the material $MnBi_2Te_4$ interfaced with black phosphorus. The researchers reached their conclusions by carefully ruling out alternative interpretations. The results establish the system as a viable setting for studying the effects of the quantum metric. —JS

Science, adf1506, this issue p. 170

QUANTUM OPTICS

Quantum nonlinear optics

In general, photons do not interact much with each other, requiring a nonlinear medium to enhance and induce an interaction. Such an interaction typically requires a large number of photons to change the refractive index of the medium. Drori *et al*. used an optically compressed and trapped sample of Rydberg atoms to induce an extreme photon-photon interaction regime in which quantum nonlinear effects can be seen. The formation of two- and three-photon vortices was demonstrated, providing a route to developing quantum nonlinear optics and multiphoton quantum logic. —ISO

Science, adh5315 this issue p. 187

WOOD ENGINEERING

CRISPR wood for sustainable fibers

Trees provide an important natural resource, but breeding for optimal wood properties is time consuming and hindered by the complexity of tree genetics and diversity. Sulis *et al*. show that CRISPR technologies can be readily deployed to enhance wood properties and augment the sustainability of forest trees (see the Perspective by Zuin Zeidler). The authors generated multiplexed genetic alterations modifying wood composition in poplar with more desirable traits for fiber pulping and lower carbon emissions. This work demonstrates that genome editing can be harnessed for breeding more efficient trees, which will provide timely opportunities for sustainable forestry and a more efficient bioeconomy. —DJ

Science, add4514, this issue p. 216; see also adi8186, p. 124

HEART DISEASE

Open the door, HOMER!

Atrial fibrillation (AFib), the most common type of heart arrhythmia, is a serious condition that can result in atrial blood clots and thromboembolic stroke. Hulsmans *et al*. performed single-cell RNA sequencing on atrial tissue from AFib patients and healthy controls to better understand how stromal and immune cells contribute to this disease. They found that recruited $CCR2^+$ $SPP1^+$ macrophages expanded in AFib patients. These cellular and transcriptomic changes were recapitulated in a mouse model of AFib that integrated hypertension, obesity, and mitral valve regurgitation (HOMER). Disrupting *Ccr2*, which coordinates inflammatory macrophage recruitment to atria, or *Spp1*, which helps to drive inflammatory fibroblast activation by macrophages, ameliorated disease burden in HOMER mice, suggesting two potential immunotherapy targets for Afib patients. —STS

Science, abq3061, this issue p. 231

LIVER FIBROSIS

Macrophages in liver fibrosis and repair

Macrophages are known to contribute to liver disease and repair, but strategies are needed to modulate their behavior to promote disease resolution. Moreno-Lanceta *et al*. showed that the E3 ubiquitin ligase RNF41 is down-regulated in macrophages during liver fibrogenesis in humans and mice. Macrophage-targeting nanoparticles bearing plasmids increased RNF41 expression in macrophages in mouse models of liver injury, resulting in improved hepatic fibrosis and regeneration. This work demonstrates the relevance of macrophage RNF41 in the development and amelioration of chronic liver disease. —CAC

Sci. Transl. Med. (2023) 10.1126/scitranslmed.abq6225

CHATGPT

Generative AI as a writing aid

Automation has historically displaced human workers in factories (e.g., automotive manufacturing) or in performing routine computational tasks. Will generative artificial intelligence (AI) tools such as ChatGPT disrupt the labor market by making educated professionals obsolete, or will these tools complement their skills and enhance productivity? Noy and Zhang examined this issue in an experiment that recruited college-educated professionals to complete incentivized writing tasks. Participants assigned to use ChatGPT were more productive, efficient, and enjoyed the tasks more. Participants with weaker skills benefited the most from ChatGPT, which carries policy implications for efforts to reduce productivity inequality through AI. —EEU

Science, adh2586, this issue p. 181

TUMOR IMMUNOLOGY

Instructing tumor cell turncoats

Cancer cells down-regulate antigen presentation to evade recognition by the immune system. Zimmermannova *et al.* reprogrammed cancer cells into tumor-derived antigen-presenting cells (tumor-APCs). Mouse or primary human cancer cells transduced with a set of transcription factors acquired a stable dendritic cell phenotype, could present endogenous tumor antigens to stimulate the effector function of $CD8^+$ T cells, and showed decreased tumorigenicity. In mouse models, intratumoral injection of tumor-APCs decreased tumor growth, increased the survival of tumor-bearing mice, and enhanced responses to immune checkpoint inhibitors. This study highlights that future immunotherapies aimed at reprogramming cancer cells in situ could combine reversing malignancy with mobilizing the cellular antigen-presenting machinery to enhance the activity of $CD8^+$ T cells toward tumors. —SHR

Sci. Immunol. (2023) 10.1126/sciimmunol.add4817

QUANTUM NETWORKS

Building a multichip quantum network

The realization of a scalable quantum entanglement network relies on the development of a scalable architecture and on incorporating hardware components that are mass manufacturable. Zheng *et al*. proposed and demonstrated a scalable architecture using photonic hybrid multiplexing. They realized a toolbox of integrated nanophotonic hybrid multiplexing devices that can be manufactured using standard industrial fabrication processes. Their chip-based quantum entanglement network could find important applications in quantum communications, metrology, and distributed quantum computing. —ISO

Science, adg9210, this issue p. 221

SUPERCONDUCTIVITY

What sets the transition temperature?

The mechanism of high-temperature superconductivity in copper oxide materials remains a mystery more than 30 years after its initial discovery. One way to shed light on this is to look for correlations between different observables in cuprate families. Wang *et al*. used electron energy-loss spectrometry to measure the charge transfer gap in a class of bismuth-based cuprates with varying numbers of copper oxide planes in the unit cell. The measured gap was negatively correlated with the maximum transition temperature, providing a design principle for further increasing the working temperature of cuprate superconductors. —JS

Science, add3672, this issue p. 227

IN OTHER JOURNALS

Edited by **Caroline Ash** *and* **Jesse Smith**

Spotted-winged fruit fly in flight over a strawberry, one of the many fruit species that this insect's larvae feeds on

EVOLUTION

A hearty eater

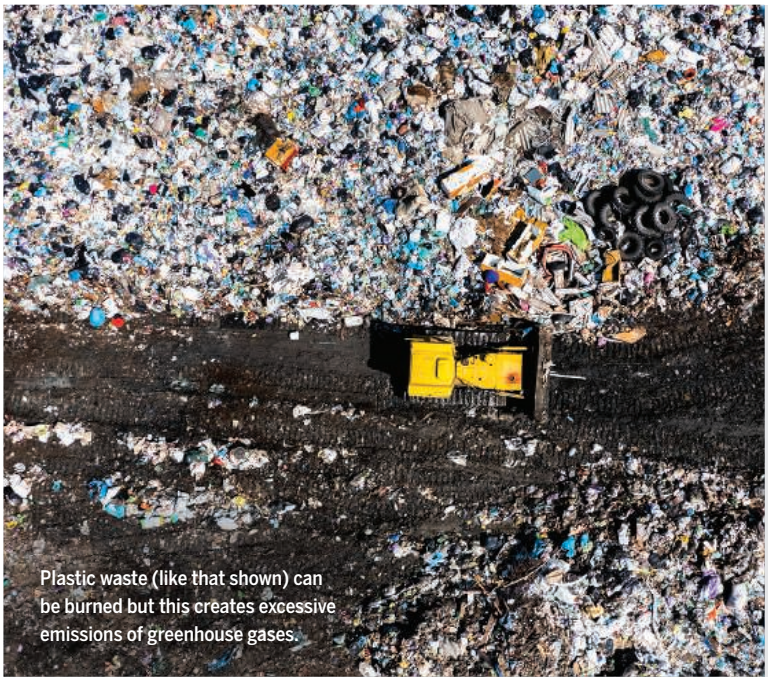
Many plant-eating insects are picky eaters and specialize on a single host plant. Some species do have generalist appetites, but whether they use the same biochemical processes as specialists to convert different foods into energy and other bioactive molecules is not well understood. Olazcuaga *et al.* fed a generalist species, the spotted wing fruit fly, on blackcurrant, cherry, cranberry, and strawberry separately and then measured the chemical compositions of the fruits and the flies. The authors found that the same metabolism was used by flies to convert fruit into energy regardless of which species they ate. Consequently, some harmful chemicals unique to a fruit were not detoxified and remained in the body. There may thus be a negative tradeoff for the advantages of being a generalist. —DJ *eLife* (2023) 10.7554/eLife.84370

NEUROSCIENCE

Unique development of the insular cortex

During fetal development, the human brain expands to form cortical lobes that fold into gyri (ridges) and sulci (grooves). Models of brain folding can explain the genesis of individual gyri and sulci. However, these models exclude the formation of the insula, a region deep inside the cortex involved in

PHOTO: NATURE PICTURE LIBRARY/ALAMY STOCK PHOTO



Plastic waste (like that shown) can be burned but this creates excessive emissions of greenhouse gases.

PLASTIC WASTE

Carbon-neutral plastic transformation

Plastic pollution has become a worldwide problem that requires integral solutions. In the past few years, plastic transformation processes such as pyrolysis, gasification, and incineration, in which high-energy products can be exploited, have become common. Kwon *et al.* now show through modeling that these processes emit large quantities of carbon dioxide and other greenhouse gases. They found that the emissions are larger than those of fossil fuel–based power systems per unit of power generated, which will become increasingly problematic as more plastic waste is used as a source of energy. The authors conclude that plastic transformation processes require effective carbon separation, capture, and storage technologies to reduce greenhouse gas emissions and comply with carbon neutrality targets. —ECF

Energy Environ. Sci. (2023) 10.1039/D3EE00969F

integrating sensory signals (interoception). Using magnetic resonance imaging and diffusion tractography, Mallela *et al.* observed that the insula has a different adult morphology and grows more slowly than other lobes and that these differences emerge early in development. In contrast to the other cortices, curved radial migratory streams of neural progenitor cells travel from a small area of the ventricular zone to reach the insular cortex. —PRS

Proc. Natl. Acad. Sci. U. S. A. (2023) 10.1073/pnas.2220200120

POLITICS

Gulf in voting

Since the 1980s, there has been a significant gender gap in US voting. Compared with men, women disproportionately vote Democratic, which is even more pronounced among unmarried voters. England *et al.* examined representative data to understand the main drivers of the gender gap and found that it stems from racial composition and racial inequality. US female voters are disproportionately Black women because Black men are disproportionately disenfranchised, incarcerated, or prematurely deceased, removing many from the electorate. Black women are also disproportionately unmarried. This discrepancy highlights the role of deeper social disparities in shaping US politics and election outcomes. —EEU

Proc. Natl. Acad. Sci. U. S. A. (2023) 10.1073/pnas.2221910120

NEUROSCIENCE

Alcohol alters action control

Previous experiences are critical for decision-making throughout life. Alcohol consumption has been shown to alter decisions, but a complete understanding of the nature of these alterations and the underlying mechanisms is lacking. Schreiner *et al.* showed that chronic alcohol administration in rodents impaired their ability to use recent experiences to guide decision-making. Hyperactivation of the premotor cortex–striatum circuit was associated with this impairment, and inhibiting the activity of this circuit restored experience-guided decision-making. Premotor regions might be potential targets to reduce alterations in executive functions associated with alcohol use disorder. —MMa

Cell Rep. (2023) 10.1016/j.celrep.2023.112675

SCIENCE EDUCATION

A DEI requirement for chemistry majors

The Amherst College chemistry community aspires for their graduates to have a critical understanding of the implications of systemic discrimination as they pertain to chemical research and education. Chung *et al.* describe the formation and ongoing efforts of a collaborative partnership among students, faculty, and staff to identify avenues for the implementation of inclusive change in their community. Specifically, the community implemented a diversity-equity-inclusion (DEI) requirement, "Chemistry in Society," which engages students in historical and modern-day challenges to equal participation and serves as the cornerstone for continuing conversations about structural inequality. Collectively, the student-staff-faculty collaboration, strengthened by intentional feedback structures, enabled inclusive departmental change and supported an ongoing feedback cycle. —MMc

J. Chem. Educ. (2023) 10.1021/acs.jchemed.2c01190

PRECISION AGRICULTURE

Minimizing fertilizer waste

Fertilizers and pesticides aid in maximizing agricultural yield, but it is preferable to administer the minimums required without wastage that can become pollutants elsewhere in the ecosystem. Aguilar Pérez *et al.* showed that the framework material ZIF-8 can be used to encapsulate MiZax-3, a synthetic mimic of the metabolite zaxinone that is known to promote plant growth. Initial tests on tomato and pearl millet seedlings showed that encapsulated MiZax-3 produced similar benefits on early-stage growth as Mizax-3 alone. Under field conditions, ZIF-8 maintained stability of MiZax-3 for 30 days. *Capsicum* plants showed enhanced yield, with an apparent synergistic effect from the ZIF-8 that may contribute zinc to aid plant growth. —MSL

Nano Lett. (2023) 10.1021/acs.nanolett.2c04506

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

Control of histone demethylation by nuclear-localized α-ketoglutarate dehydrogenase

Fei Huang†, Xiao Luo†, Yang Ou†, Zhaoxu Gao†, Qiming Tang, Zhenzhen Chu, Xinguang Zhu, Yuehui He*

INTRODUCTION: Methylations on nucleosomal histones play fundamental roles in regulating eukaryotic transcription and other chromatin-based processes. Jumonji C domain–containing histone demethylases (JMJs) are found in most eukaryotes, including animals, plants, and fungi, and dynamically control histone methylations. JMJs thus play a general role in regulating gene expression. JMJs are α-ketoglutarate (α-KG)/Fe(II)–dependent dioxygenases that use the cosubstrates α-KG and oxygen to remove methyl groups from lysine residues by oxidation.

RATIONALE: The histone demethylation activity of certain JMJs can be regulated by several mechanisms in plants and/or animals, including posttranslational modifications of some JMJs and regulation by several JMJ-associated proteins, metabolites, and the cosubstrate oxygen. It is currently unknown how nuclear JMJ activity may be regulated by other enzymes that metabolize α-KG.

RESULTS: We constructed several loss-of-function *Arabidopsis* mutants in α-ketoglutarate dehydrogenase (KGDH), and found that KGDH not only played a primary role in the control of total cellular level of α-KG, TCA cycle flux, and cellular respiration, but also regulated various aspects of growth and development. Mutants with defective KGDH exhibited an accumulation of the potent antioxidants anthocyanins, a substantial delay in the developmental transition to flowering, and a slowdown of shoot development. Next, using confocal fluorescent imaging, we found that a small portion of KGDH entered the nucleus. This process was regulated by exposure to light. We mutagenized a nuclear localization signal in a KGDH subunit, E2, and found that mitochondria-localized KGDH was unable to rescue the growth and development abnormalities in an *e2* mutant.

Biochemical measurements of various histone methylation marks revealed that global histone demethylation appeared to be inhibited by nuclear KGDH. We hypothesized that KGDH may couple with JMJs to regulate histone methylations. Indeed, further protein-protein interaction experiments revealed that nuclear KGDH directly interacted with various JMJs including demethylases of histone 3 lysine 4, histone 3 lysine 9, and histone 3 lysine 27.

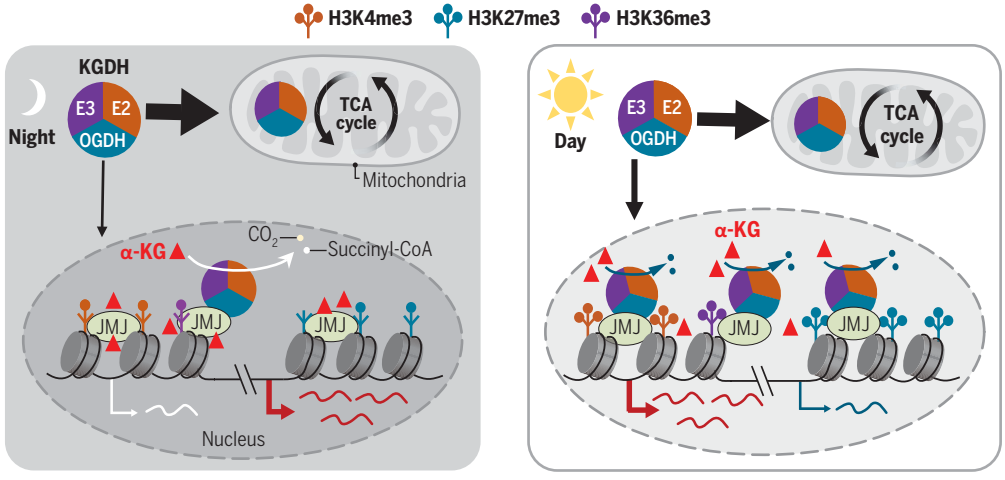
We explored the function of nuclear KGDH in controlling genome-wide histone methylation and gene expression in *Arabidopsis* using RNA sequencing and chromatin immunoprecipitation coupled with DNA sequencing. KGDH was found to occupy genomic regions at thousands of loci, correlating with histone methylation regions. Transcriptomic profiling revealed that thousands of genes, particularly environment-responsive genes, were misregulated in mutants with defective KGDHs. Biochemical analysis uncovered that nuclear KGDH catalyzed oxidative decarboxylation of α-KG to reduce the local availability of α-KG to the JMJs coupled to KGDH. Reduced α-KG resulted in inhibition of histone demethylation by JMJs. We further engineered a KGDH subunit to enhance its nuclear targeting, which led to a further inhibition of histone demethylation by JMJs. Moreover, we found that nuclear KGDH-JMJ assemblies can activate or repress the expression of target loci, depending on the readout of local histone methylation marks.

Through protein-protein interaction experiments, we found that the evolutionarily conserved carboxyl-terminal regions of a KGDH subunit from both human and mouse interacted with mammalian JMJs. Thus, KGDH association with JMJs is conserved in mammalian cells and potentially in other eukaryotic organisms.

CONCLUSION: Here, we have uncovered that the TCA cycle–associated enzyme KGDH enters the nucleus, and that this process is regulated by light in plants. Nuclear KGDH interacts with various JMJs, catalyzes oxidative decarboxylation of α-KG, and thus inhibits α-KG–dependent histone demethylation by JMJs to regulate genome-wide gene expression in plants. The association of nuclear KGDH with JMJs is also found in mammalian cells, suggesting that KGDH may function together with JMJs to control histone demethylations and gene expression in mammals in addition to plants. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: yhhe@pku.edu.cn
†These authors contributed equally to this work.
Cite this article as F. Huang *et al.*, *Science* **381**, eadf8822 (2023). DOI: 10.1126/science.adf8822

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adf8822



Light promotes nuclear targeting of the TCA cycle–associated enzyme KGDH to regulate histone demethylation by JMJs in plants. Most KGDH subunits, including OGDH1, E2, and E3, are targeted to mitochondria to form a holoenzyme for α-KG decarboxylation in the TCA cycle. Light exposure promotes a small portion of KGDH entering into the nucleus. There, KGDH occupies thousands of loci and interacts with various JMJ histone demethylases to inhibit α-KG–dependent histone demethylations by JMJs. Inhibition of histone demethylation occurs by limiting the local availability of α-KG, and consequently controls gene expression.

RESEARCH ARTICLE

PLANT SCIENCE

Control of histone demethylation by nuclear-localized α-ketoglutarate dehydrogenase

Fei Huang[1,2]†, Xiao Luo[2,3]†, Yang Ou[1,2]†, Zhaoxu Gao[1]†, Qiming Tang[4], Zhenzhen Chu[2], Xinguang Zhu[4], Yuehui He[1,2,3]*

Methylations on nucleosomal histones play fundamental roles in regulating eukaryotic transcription. Jumonji C domain–containing histone demethylases (JMJs) dynamically control the level of histone methylations. However, how JMJ activity is generally regulated is unknown. We found that the tricarboxylic acid cycle–associated enzyme α-ketoglutarate (α-KG) dehydrogenase (KGDH) entered the nucleus, where it interacted with various JMJs to regulate α-KG–dependent histone demethylations by JMJs, and thus controlled genome-wide gene expression in plants. We show that nuclear targeting is regulated by environmental signals and that KGDH is enriched at thousands of loci in *Arabidopsis thaliana*. Chromatin-bound KGDH catalyzes α-KG decarboxylation and thus may limit its local availability to KGDH-coupled JMJs, inhibiting histone demethylation. Thus, our results uncover a regulatory mechanism for histone demethylations by JMJs.

Nuclear DNA wraps around histone proteins to form nucleosomes that constitute chromatin. Methylations on histones regulate chromatin structure, transcription, and other chromatin-based processes in eukaryotes (*1*, *2*). Commonly methylated sites on histones include lysine 4 (K4), K9, K27, K36, and K79 on histone 3 (H3) and K20 on histone 4 (H4) (*2*). Histone lysine methyltransferases (KMTs) catalyze the methylation of lysine residues on histones, whereas histone lysine demethylases (KDMs) remove the methyl groups. Histone methylation is dynamically altered by KMTs and KDMs in response to endogenous and external inputs (*1*, *3*).

KDMs are composed of two evolutionarily conserved families, lysine-specific demethylases (LSDs), in the KDM1 subfamily, and Jumonji C (JmjC) domain–containing histone demethylases (JMJs, JHDMs, or JMJDs), a large family composed of a few subfamilies. Seven subfamilies (KDM2 to KDM7) have been found in mammals (*1*, *4*). KDM1s catalyze flavin-dependent oxidative demethylation of mono- and dimethyl lysine residues including H3K4 and H3K9 (*1*). JMJs are α-ketoglutarate (α-KG)/Fe(II)–dependent dioxygenases, and the catalytic JmjC domains use the cosubstrates α-KG and oxygen to remove the methyl groups from the lysine residues by oxidation (*4*, *5*). JMJs are found in most eukaryotes, including animals, plants, and fungi, and play a general role in regulating gene expression (*3*, *6*). In *Arabidopsis thaliana*, there are 21 potential JmjC domain–containing histone demethylases (*7*, *8*). Among them, the KDM5 subfamily members, including JMJ14 to JMJ18, function as H3K4 demethylases that preferentially demethylate trimethyl H3K4. By contrast, mono- and dimethyl H3K4 are typically demethylated by the KDM1 family members, including three *Arabidopsis* homologs of the human lysine–specific demethylase 1 (LSD1) (*9*–*11*). Six members of the KDM3 subfamily, including JMJ25/IBM1 (INCREASE IN BONSAI METHYLATION 1) and JMJ26 to JMJ29, function as H3K9 demethylases (*11*, *12*). In addition, homologs of the mammalian KDM4 family (H3K9/H3K36 demethylases) have been found in *Arabidopsis*, including JMJ11/ELF6 (EARLY FLOWERING 6), JMJ12/REF6 (RELATIVE OF EARLY FLOWERING 6), and JMJ13, but these JMJs demethylate H3K27 with a substrate preference for trimethyl K27 (*13*–*15*). Members of the mammalian KDM2 (H3K36 demethylases) and KDM6 (H3K27 demethylases) subfamilies have not been found in land plants (*6*). Several JmjC domain enzymes or α-KG/Fe(II)–dependent dioxygenases may mediate H3K36 demethylation in *Arabidopsis* (*16*, *17*), but their activities have yet to be validated.

JMJ-mediated histone demethylation requires α-ketoglutarate, an obligatory substrate generated primarily by the tricarboxylic acid (TCA) cycle in the mitochondria matrix (*5*, *18*). α-KG is produced by isocitrate dehydrogenase (IDH) through oxidative decarboxylation of isocitrate, and is subsequently decarboxylated oxidatively by α-ketoglutarate dehydrogenase (α-KGDH) to generate succinyl-CoA and CO_2 (*19*, *20*). These two reactions are rate-limiting steps in the TCA cycle (*19*, *20*). The KGDH enzyme complex is composed of three subunits including oxoglutarate dehydrogenase (OGDH, E1 subunit), dihydrolipoamide succinyltransferase (DLST, E2 subunit), and dihydrolipoamide dehydrogenase (DLD, E3 subunit) (*21*, *22*). OGDH, E2, and E3 catalyze α-KG decarboxylation, the formation of succinyl-CoA, and dihydrolipoamide dehydrogenation, respectively (*22*). In *Arabidopsis*, the E1 subunit OGDH is encoded by two genes, *OGDH1* and *OGDH2* (*22*). Putative *OGDH* loss-of-function mutants (loss of both *OGDH1* and *OGDH2* function) have been reported in which levels of total cellular α-KG remained unchanged compared with wild type (WT) (*22*). As a result, the role of KGDH in the control of the level of total cellular α-KG is still unclear. In addition, these putative double mutants exhibited normal growth and development with only slight phenotypic changes (*22*). Biological functions for KGDHs other than their role in the TCA cycle are not clear.

Cytosolic α-KG molecules diffuse into the nucleus through nuclear pores and serve as an obligatory substrate for histone demethylation by JMJs (*23*, *24*). α-KG is a rate-limiting substrate for the family of α-KG/Fe(II)–dependent dioxygenase (*19*, *25*, *26*). Michaelis constant (K_m) values for α-KG of the JMJs examined so far including human KDM4, KDM5, and KDM6 family enzymes are in the range of 6 to 50 μM, and the local physiological concentrations of α-KG for JMJs are estimated in this range (*27*–*29*). Therefore, JMJ activities may be sensitive to fluctuations in nuclear α-KG concentrations. In mouse embryonic stem cells, a reduction in the level of intracellular α-KG reduces the rate of histone demethylations by JMJs (*25*). The local availability or concentration of α-KG may regulate the activity of nuclear JMJs. In this study, using the model plant *Arabidopsis*, we found that the TCA cycle–associated enzyme KGDH controls the nuclear availability of α-KG to regulate histone demethylations by JMJs.

KGDH controls α-KG levels and regulates growth and development in *Arabidopsis*

To explore the metabolic and biological functions of KGDH, we constructed loss-of-function mutants of the E1/OGDH subunits of KGDH. Two putative *ogdh1 ogdh2* double mutants with a normal phenotype were previously reported (*22*). However, we observed that in one described *ogdh2* allele (Salk_055824), the putative mutagen, a transferred DNA (T-DNA), is inserted downstream of *OGDH2* (fig. S1, A, D, and E), whereas another described *ogdh2* allele (Salk_122458) carries multiple segregating T-DNAs (https://Arabidopsis.info/). We first identified two loss-of-function *ogdh1* mutants in which

[1]National Key Laboratory of Wheat Improvement, Peking-Tsinghua Center for Life Sciences, School of Advanced Agricultural Sciences, Peking University, Beijing 100871, China. [2]Shanghai Center for Plant Stress Biology, CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences (CAS), Shanghai 201602, China. [3]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong 261325, China. [4]National Key Laboratory for Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, CAS, Shanghai, 200032, China.
*Corresponding author. Email: yhhe@pku.edu.cn
†These authors contributed equally to this work.

the full-length transcripts of *OGDH1* were eliminated by insertional T-DNAs (fig. S1, A to C). Next, we generated two *ogdh2* mutants using CRISPR/CRISPR-associated protein 9 (Cas9). A 9-bp deletion in *ogdh2-1* removed three amino acid residues from the OGDH2 protein, and a single–base pair insertion frameshifted *ogdh2-2* (fig. S1D).

We found that the single mutants of *ogdh1* and *ogdh2* flowered moderately later than WT, based on the total number of primary leaves formed before flowering (Fig. 1C and fig. S1F). We further found that the double mutants of *ogdh1 ogdh2* flowered very late even under inductive long-day growth conditions (Fig. 1, A and C, and fig S1G), and plastochron growth in the *ogdh1 ogdh2* mutants were much slower than in the WT (fig. S1H). These factors gave rise to greatly prolonged "days to flowering" (fig. S1G). The late-flowering phenotype of *ogdh1 ogdh2* was rescued by the introduction of WT genomic *OGDH1* or *OGDH2* fragments (fig. S2A), suggesting functional redundancy in promoting floral transition.

To further determine whether OGDH functions as part of KGDH to control floral transition, we examined loss-of-function mutants of E2 subunits. The *Arabidopsis* E2 subunit is encoded by two genes (*30*), *E2a* (*At5G55070*) and *E2b* (*At4G26910*), with very similar protein sequences (figs. S3A and S4, A and C). We identified loss-of-function mutants in each gene, *e2a-1* and *e2b-1* (fig. S3B). Both *e2a* and *e2b* mutants, like *ogdh1* and *ogdh2*, flowered later than WT (Fig. 1E and fig. S3D). An *e2a* allele with a T-DNA insertion in its 3′ untranslated region was recently found to flower slightly later (*30*). We failed to obtain an *e2a e2b* double mutant through genetic crossing, likely due to an essential function of E2/KGDH in cellular respiration.

In addition to the developmental abnormalities, we further observed that the regions encompassing the shoot apical meristem in *ogdh1 ogdh2* turned light purple (Fig. 1B), indicating that there was an accumulation of anthocyanins. Anthocyanins function as potent antioxidants to scavenge reactive oxygen species and other free radicals in vegetative tissues during biotic and abiotic stresses (*31*, *32*). We measured the levels of total anthocyanins in the seedlings of WT, *ogdh1*, *ogdh2*, and *ogdh1 ogdh2*, and found that in the *ogdh1 ogdh2* double mutants, anthocyanins accumulated to higher levels than in the WT or single mutants (Fig. 1D and fig. S2C), revealing that *OGDH1* and *OGDH2* function redundantly to inhibit anthocyanin biosynthesis under normal growth conditions. Moreover, we found that the levels of anthocyanins were higher in both *e2a* and *e2b* mutants relative to WT (Fig. 1F). Thus, *OGDH* and *E2* genes act to prevent biosynthesis of anthocyanins in *Arabidopsis* seedlings grown under normal conditions.

Next, to determine the extent of perturbation in cellular primary metabolism upon loss of *OGDH* function, we first measured total cellular α-KG levels in whole seedlings of WT and *ogdh1-2 ogdh2-1* (hereafter referred to as *ogdh1/2*), and found increased α-KG levels (Fig. 1G). This shows that OGDH/KGDH plays a primary role in controlling the level of cellular α-KG by catalyzing its oxidative decarboxylation. Furthermore, we measured the cellular levels of TCA cycle intermediates and found that succinate (the product of α-KG decarboxylation) and malate were greatly decreased in *ogdh1/2* seedlings compared with WT (fig. S5A), suggesting a reduction in the TCA cycle flux upon a functional loss of KGDH. We also found that cellular levels of most Calvin cycle and glycolysis intermediates were increased in *ogdh1/2* compared with WT. For example, there was an ~3-fold increase in sucrose (fig. S5A). This indicates that there may be a slowdown of glycolysis caused by a reduction in TCA cycle flux upon a functional loss of the rate-limiting OGDH/KGDH, which is consistent with the slowdown of shoot growth and development (slower rates of leaf initiation) in the *ogdh* mutants.

We found that succinate, fumarate, and malate were all decreased in *e2a* seedlings compared with WT (fig. S5B), indicating a reduction in the TCA cycle flux. Unlike in *ogdh1/2*, the level of total cellular α-KG in *e2a* was similar to that in WT (Fig. 1H). In KGDH-catalyzed oxidative decarboxylation of α-KG, the E1/OGDH subunit catalyzes decarboxylation of α-KG to produce succinyl–thiamine pyrophosphate that is later converted into succinyl-CoA by the E2 subunit (*33*). OGDH can catalyze α-KG decarboxylation in the absence of E2 (*33*); therefore, this subunit remains active in the *e2a* seedling, accounting for the normal turnover of α-KG in *e2a*. These results together show that a disruption of KGDH function causes a reduction in the TCA cycle flux and a slowdown of cellular respiration in *Arabidopsis* seedlings.

KGDH enters into the *Arabidopsis* nucleus

Because α-KG is an obligatory substrate for histone demethylation by JMJs (*5*), elevation of total cellular α-KG in *ogdh1/2* might give rise to misregulation of various genes and consequent growth and development abnormalities. However, this is unlikely, because the *e2* mutants exhibit similar phenotypes as *ogdh1/2* but have a normal level of total cellular α-KG as that in WT. This rules out the possibility of the level of total cellular α-KG in *ogdh1/2* exerts a critical role to regulate gene expression. Alternatively, we reasoned that a minor portion of the TCA cycle–associated KGDH enzymes may enter into the nucleus to directly regulate gene expression.

It has been reported that a minor portion of a subset of mitochondria TCA cycle enzymes can move to the nuclei of several types of mammalian cells (*21*, *34*), but it is unknown whether this occurs in other eukaryotic cells. Although the enzymatic activity–associated domains in OGDH and E2 are evolutionarily conserved, the rest of the protein sequences are not well conserved between *Arabidopsis* and mammals (figs. S1A and S4, A and C). Therefore, we investigated whether KGDH may enter the nucleus in *Arabidopsis*. Using transgenic seedlings expressing functional OGDH1:GFP, OGDH2:GFP or E2a:GFP (translational fusions driven by respective native promoters), we found that most of the OGDH and E2 proteins localized to mitochondria (labeled by Mitotracker red staining), but a minor portion of OGDH1, OGDH2, and E2a localized to nuclei (Fig. 2, A to D, and fig. S6, A and C). In addition, we found that a minor portion of E2b and E3a (AT3G17240; tagged by HA) also localized to nuclei (Fig. 2, E and F). We further found OGDH- and E2-dependent KGDH activity in the nucleus (fig. S6D). Thus, a minor portion of the KGDH holoenzymes are located in the nuclei of *Arabidopsis* seedlings.

To determine how the KGDH holoenzyme enters the nucleus, we examined the sequences of OGDH, E2, and E3 subunits using *NLStradamus* (*35*), and found that there was a nuclear localization signal (NLS) in the E2 subunits, but not in the OGDH1, OGDH2, or E3 subunits (fig. S4A), indicating that KGDH nuclear targeting may depend on E2. In addition, we observed that E2s from other eukaryotic organisms such as green algae, moss, yeast, and fly also bear a predicted NLS (fig. S4C). We mutated R236, R237, R242, and R244 in the putative NLS (fig. S4A) and found that these mutations (in mtE2a-1) prevented nuclear accumulation of E2a, but not its localization in mitochondria (Fig. 2B and fig. S6, A and B). Similarly, simultaneous mutations of K221, S223, K232, and R234 in the NLS of E2a (mtE2a-2) prevented its nuclear localization (fig. S6A). Furthermore, we found that in the *e2a* mutant, the nuclear abundance of OGDH1 was greatly reduced (Fig. 2I), which was also reduced in *e2b* (fig. S6E).

We further found that the mitochondrion-localized E2a:GFP proteins with a mutated NLS were not able to complement the *e2a* mutant, whereas the E2a:GFP proteins localized in both nucleus and mitochondrion fully rescued the abnormalities of growth and development in *e2a* (fig. S3, E to H). Thus, nuclear KGDH is essential for its biological functions.

Light promotes nuclear KGDH targeting

Because light is the single most important environmental factor for plant growth and development (*36*), we explored the light regulation of KGDH moving into the nucleus in early *Arabidopsis* seedlings. We found that light exposure for 6 hours induced an accumulation

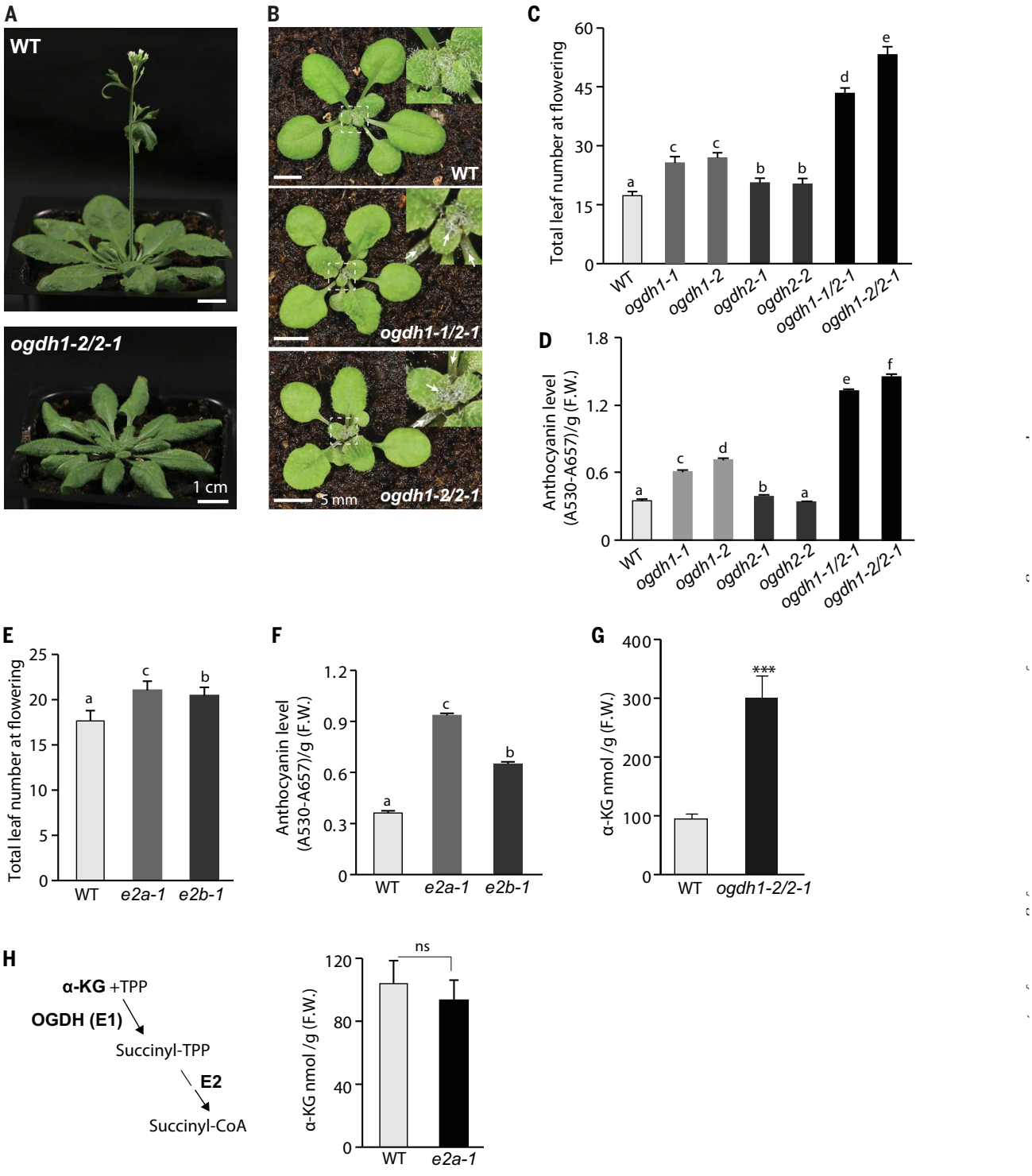


Fig. 1. Phenotypes in *Arabidopsis* KGDH-related mutants. (**A**) *ogdh1 ogdh2* double mutants grown in inductive LDs flowered extremely late. Scale bars, 1 cm. (**B**) *ogdh1 ogdh2* double mutants accumulated anthocyanins in the region encompassing the shoot apical meristem (brown-purple tissues indicated by arrows). Scale bars, 5 mm. (**C**) Flowering times of *ogdh* mutants grown in LDs. The total number of primary leaves formed before flowering in 15 plants per line was counted. (**D**) Anthocyanin content in *ogdh* seedlings. Relative content was calculated by (A530 − A657)/g (fresh weight), and three biological replicates were performed for each genotype. (**E**) Flowering times of *e2a* and *e2b* mutants grown in LDs. Fifteen plants per line were scored. (**F**) Anthocyanin content in WT, *e2a*, and *e2b* seedlings. Three biological replicates were performed for each genotype. (**G**) α-KG levels in WT and *ogdh1 ogdh2* seedlings. (**H**) α-KG levels in WT and *e2a* seedlings. Left, diagram illustrating the reactions catalyzed by OGDH/E1 and E2 in the KGDH holoenzyme. TPP, thiamine pyrophosphate. In (C) to (H), values are means ± SD [(C) and (E)] or SD of three biological replicates [(D) and (F) to (H)]. Letters in (C) to (F) mark statistically significant differences (one-way ANOVA, $P < 0.01$). In (G) and (H), ***$P < 0.001$ and ns, significant (two-tailed Student's *t* test).

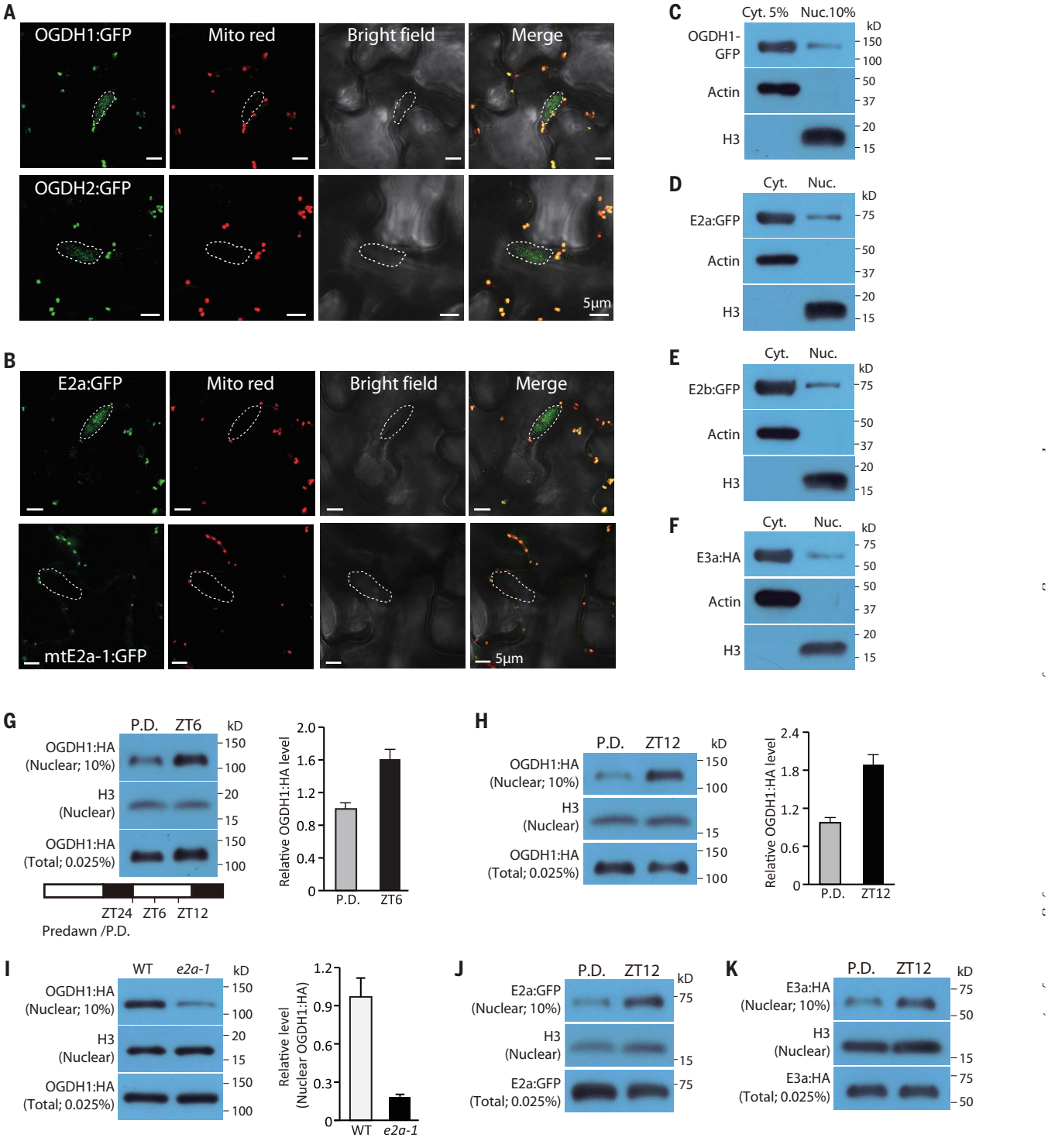


Fig. 2. Light promotes the KGDH holoenzyme entering the nucleus. (**A**) Subcellular localization of OGDH1:GFP and OGDH2:GFP in leaf epidermal cells. Mitochondria are labeled by Mitotracker red staining. Scale bars, 5 μm. (**B**) Subcellular localization of E2a:GFP and mtE2a:GFP in leaf epidermal cells. In mtE2a-1:GFP, four arginine residues in the putative NLS were mutated to alanines. Scale bars, 5 μm. (**C** to **F**) Analysis of subcellular localization of KGDH subunits by immunoblotting. Cytoplasmic (cyt.) and nuclear (nuc.) proteins were extracted separately from the seedlings expressing OGDH1:GFP (C), E2a:GFP (D), E2b:GFP (E), or E3:HA (F). Actin and H3 served as cytoplasmic fraction and nuclear fraction markers, respectively. A total of 5% of cytoplasmic and 10% of nuclear proteins were loaded for Western blotting. Seedlings grown in LDs were harvested at ZT6. (**G** and **H**) Examination of nuclear OGDH1:HA abundance in young seedlings at predawn, ZT6 (G), and ZT12 (H) by immunoblotting. Right: relative levels of nuclear OGDH1:HA at the indicated time points. Values are means ± SD of six biological replicates. (**I**) Dependence of OGDH1 nuclear localization on E2. Immunoblotting was conducted to analyze fractioned proteins from a functional *OGDH1:HA* line (fig. S2A) and *OGDH1:HA e2a*. Right: relative levels of nuclear OGDH1:HA in WT and *e2a* seedlings (harvested at ZT6). Values are means ± SD of six biological replicates. (**J** and **K**) Analysis of nuclear E2a (J) and E3a (K) levels at predawn and ZT12 by immunoblotting.

of the OGDH1 proteins in nuclei in young seedlings, and that a 12-hour exposure of light induced more OGDH1 proteins to be localized in nuclei (Fig. 2, G and H). In addition, we observed elevated nuclear accumulation of the other two subunits of KGDH, including E2a and E3a, upon light exposure (Fig. 2, J and K).

Nuclear KGDH functions to inhibit various histone demethylations

The catalytic activities of dioxygenases such as JMJs are dependent on the cosubstrate α-KG (*4*). KGDH consumes α-KG and may function to inhibit histone demethylations by JMJs. Therefore, we investigated whether loss of OGDH/KGDH function affects histone methylations. We found that H3K4me3, H3K9me2, H3K27me3, and H3K36me3 levels were all greatly decreased in the *ogdh1/2* seedlings relative to WT, whereas in *ogdh1* and *ogdh2* single mutants, the levels of these marks were moderately reduced or similar compared with WT (Fig. 3, A and B). In addition, we found that loss of *OGDH1/2* function had no effect on the level of total histones modified with H3K4me1, H3K4me2, or H3K27me2 (Fig. 3, A and B). These results suggest that *OGDH1* and *OGDH2* redundantly inhibit global histone demethylations of H3K4me3, H3K9me2, H3K27me3, and H3K36me3.

H3K4me3, H3K9me2, and H3K27me3 are known to be demethylated only by JMJs in *Arabidopsis*, whereas demethylation of H3K4me1 and H3K4me2 is catalyzed by the *Arabidopsis* homologs of the human LSD1 H3K4 demethylase that does not use α-KG as a substrate (*9*–*11*, *13*). Specifically, the *Arabidopsis* H3K27 demethylases, including ELF6, REF6, and JMJ13, preferentially demethylate H3K27me3; KDM5 family members such as JMJ14 preferentially demethylate H3K4me3; and KDM3 subfamily members such as IBM1 demethylate H3K9me2 (*9*, *11*, *13*, *15*). In addition, H3K36me3 seems to be demethylated by JMJs (*16*). The decreases of H3K4me3, H3K9me2, H3K27me3, and H3K36me3, but not H3K4me1, H3K4me2, or H3K27me2, on total histones upon loss of OGDH/KGDH function suggest that KGDH functions to inhibit α-KG–dependent histone demethylations catalyzed by JMJs.

To assess genome-wide changes of H3K27me3 upon loss of *OGDH* function, we performed chromatin immunoprecipitations with anti-H3K27me3, followed by DNA sequencing (ChIP-seq), using WT and *ogdh1/2* seedlings. Consistent with previous findings (*37*, *38*), H3K27me3 occurred mainly in gene bodies in the *Arabidopsis* genome (Fig. 3C and fig. S7A). Furthermore, we observed that loss of *OGDH1/2* function caused an overall reduction of H3K27me3 in gene bodies (Fig. 3C and fig. S7B), corroborating the strong reduction of H3K27me3 on total histones in *ogdh1/2* relative to WT (Fig. 3B).

We further examined the levels of histone methylations on total histones extracted from the *e2a* seedlings and found that, like in *ogdh1/2*, the levels of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 were strongly decreased upon loss of *E2a* function (Fig. 3D). Similarly, the levels of these histone marks were decreased in the *e2b* seedlings (fig. S6F). These results further show that KGDH acts to inhibit histone demethylations by JMJs. Because the level of total cellular α-KG remains unchanged in *e2a* relative to WT (Fig. 1H), the reduced levels of methylated H3K4, H3K9, H3K27, and H3K36 in *e2a* are obviously not caused by a fluctuation in the level of total cellular α-KG. Given the E2-dependent nuclear targeting of KGDH and the biochemical role of KGDH in the oxidative decarboxylation of α-KG, we reasoned that an increase in the level of nuclear α-KG in *e2a* (and *e2b* and *ogdh1/2*) may have caused the reductions in methylated histones through elevated availability of local α-KG to JMJs. Indeed, we found that the nuclear E2a, but not mitochondrial E2a, was responsible for the inhibition of JMJ-mediated histone demethylations (Fig. 3D). Together, these results suggest that nuclear KGDH catalyzes oxidative decarboxylation of α-KG to reduce its local concentration and thus inhibit JMJ activities.

Nuclear KGDH regulates environment-responsive gene expression

To investigate the role of *OGDH1/2* and *E2a* in transcriptome regulation, we performed RNA sequencing (RNA-seq) and found that 1008 and 587 genes were up-regulated, whereas 1572 and 791 genes were down-regulated in *ogdh1/2* and *e2a* relative to WT, respectively [cutoff value: 1.8-fold; false discovery rate (FDR) < 0.05)] (Fig. 3, E and F; fig. S8, A to C; and table S1). A total of 78% of the up-regulated genes (422 of 544) and 89% of the down-regulated genes in *e2a* (702 of 791) were also up- or down-regulated in *ogdh1/2*, respectively (Fig. 3, E and F, and fig. S8D), consistent with the idea that OGDH (E1) and E2 function as part of the KGDH holoenzyme to regulate nuclear gene expression.

We examined the chromatin features of *KGDH*-regulated loci by comparing them with H3K27me3- or H3K4me3-modified loci in *Arabidopsis* seedlings, because H3K4me3 and H3K27me3 are linked with transcriptional activation and repression, respectively (*3*, *39*). A total of 257 up-regulated genes in the *ogdh1/2* seedlings were H3K27me3-bearing loci, and 635 down-regulated genes in *ogdh1/2* were H3K4me3-bearing loci (Fig. 3, E and F, and table S2), revealing significant overlap. Furthermore, we found that in most of the 257 up-regulated genes in the *ogdh1/2* seedlings, the levels of repressive H3K27me3 were reduced compared with WT, particularly in the gene-body regions where H3K27me3 is enriched (Fig. 3H), suggesting that the nuclear OGDH/KGDH inhibits JMJ-mediated demethylation of H3K27me3 to repress the expression of these genes.

To understand the biological functions of *OGDH*-regulated genes, Gene Ontology (GO) enrichment analysis was conducted. Most of top over-represented or enriched GO terms of biological process were related to responses to environmental stimuli (biotic and abiotic stimuli) and stress hormones including abscisic acid and jasmonic acid (fig. S8E). The GO term "response to red and far red light" was enriched in the *OGDH*/*KGDH* regulon (fig. S8E), consistent with light promoting KGDH movement into the nucleus (Fig. 2, G and H). Furthermore, the expression of genes encoding biosynthesis of anthocyanins (stress marker metabolites) was enriched in the GO analysis (fig. S8E), consistent with anthocyanin accumulation upon functional loss of KGDH in the *ogdh1/2* and *e2* mutants (Fig. 1, D and F).

Next, we investigated whether the regulatory genes in the floral transition or anthocyanin biosynthesis may be misregulated upon functional loss of KGDH. The cold-responsive MADS-box transcription factor *FLOWERING LOCUS C* (*FLC*) is a central floral repressor (*40*). We found that *FLC* expression was derepressed in both *ogdh1/2* and *e2a* (with a >4-fold increase in *ogdh1/2*) (Fig. 3, G and I). The late-flowering phenotype in *ogdh1/2* was largely rescued by an *flc* mutation (fig. S2D). Furthermore, we found that the expression of the master regulator of anthocyanin biosynthesis, *PRODUCTION OF ANTHOCYANIN PIGMENT 1* (*PAP1*) (*31*, *32*), was greatly derepressed in both *ogdh1/2* and *e2a* (~20-fold increase in *ogdh1/2*) (Fig. 3, G and I). The stress-responsive MYB-domain transcription factor PAP1 activates the expression of structural genes in anthocyanin biosynthesis (*31*, *32*). We found that a *pap1* mutation largely eliminated the anthocyanin accumulation caused by loss of *OGDH1/2* function (Fig. 3J and fig. S8F), revealing that *OGDH1* and *OGDH2* repress *PAP1* expression to inhibit anthocyanin biosynthesis.

In an *e2a* line expressing only mitochondria-localized E2a, we found that the derepression of neither *FLC* nor *PAP1* in *e2a* was rescued (Fig. 3I). This confirms that the derepression of both genes is caused by the loss of nuclear E2/KGDH activity. We thus conclude that the nuclear KGDH, but not the mitochondria KGDH alone, is essential for transcriptome regulation.

OGDH interacts with JMJs in the nucleus

We found that nuclear KGDH functions to inhibit histone demethylations by JMJs. To address how nuclear KDGH may inhibit JMJ activities, we first investigated whether OGDH/KGDH may interact with individual JMJs. Using yeast two-hybrid analysis, we found that the C-terminal regions of OGDH1 and OGDH2 (composed of a conserved pyrimidine-binding

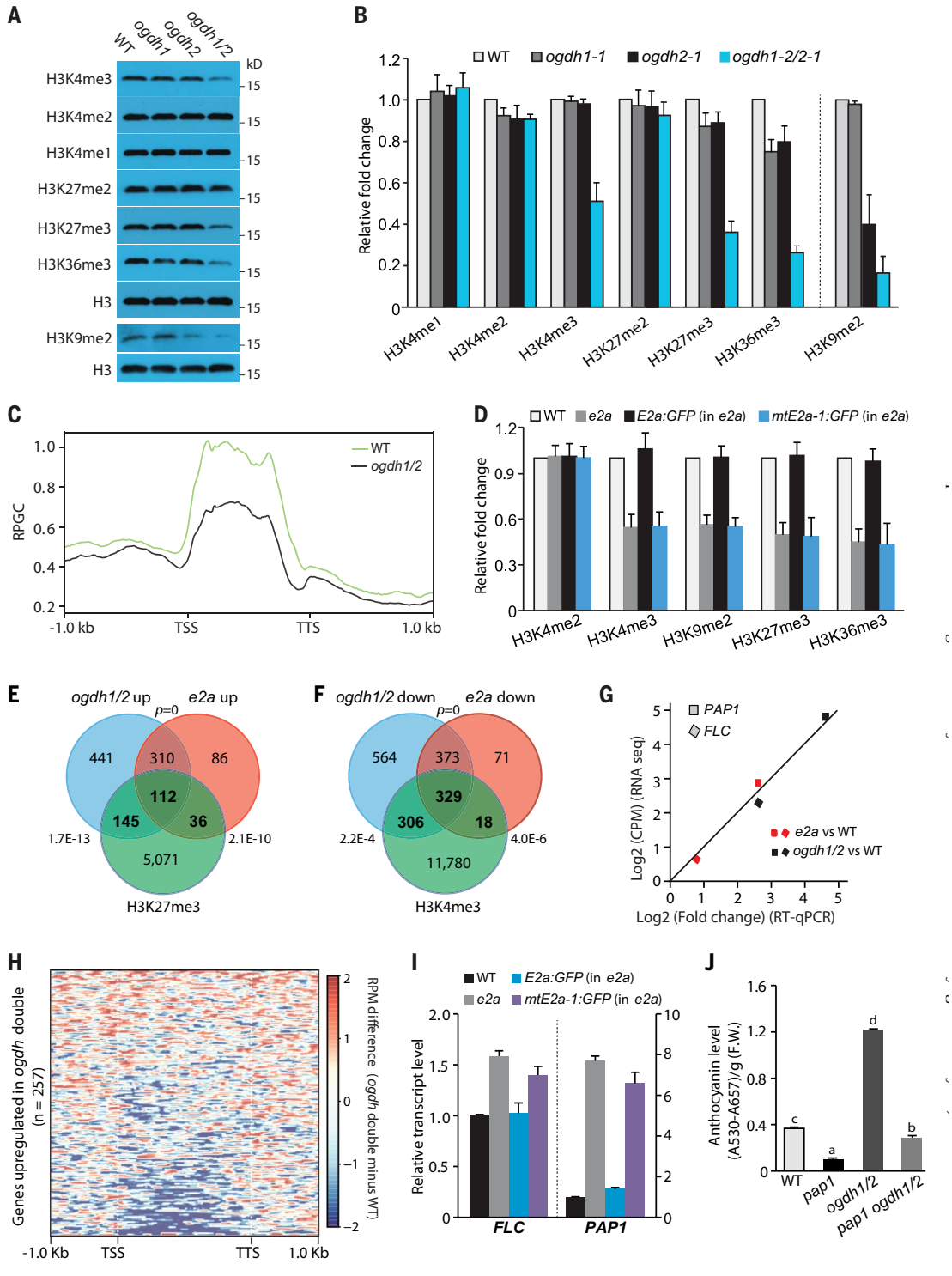


Fig. 3. Nuclear KGDH functions to inhibit histone demethylations and regulate global gene expression. (**A** and **B**) Levels of H3 methylated at various lysine residues in the indicated *Arabidopsis* seedlings as revealed by immunoblotting. Total histones extracted from the seedlings were analyzed. Shown in (A) is a set of representative blots from three biological repeats. Relative levels of H3 methylated at K4, K9, K27, or K36 are shown in (B). Values are means ± SD of three biological replicates. (**C**) Metagene plots of H3K27me3 enrichment in WT and *ogdh1/2* seedlings. RPGC, reads per genomic content; TTS, transcription termination site. (**D**) Levels of H3 methylated at K4, K9, K27, or K36 in *e2a*-related seedlings, as revealed by immunoblotting. Relative levels are shown, and values are means ± SD of five biological replicates. (**E** and **F**) Venn diagrams of the genes up-regulated (E) or down-regulated (F) in *ogdh1/2* and/or *e2a* seedlings overlapped with H3K27me3- or H3K4me3-occupied loci. H3K4me3-modified genes (at a seedling stage) are extracted from GSE185121. Overlapping *P* values (Fisher's exact test) are indicated. (**G**) Validation of RNA-seq data by qPCR. Plotted are fold changes of the transcript levels of *FLC* and *PAP1* in *ogdh1/2* and *e2a* (versus WT) quantified by RNA-seq or RT-qPCR. Values are means of three biological replicates. (**H**) Heatmap showing H3K27me3 changes at the 257 up-regulated loci in *ogdh1/2* (versus WT). Blue color indicates a reduction of H3K27me3 in *ogdh1/2* relative to WT. RPM, reads per million. (**I**) Quantification of the transcript levels of *FLC* and *PAP1* in the indicated seedlings. Transcript levels were first normalized to the constitutively expressed *TUBULIN2* (*TUB2*), and relative expression to WT is shown. Values are means ± SD of three biological replicates. (**J**) Anthocyanin content in the seedlings of WT, *pap1*, *ogdh1/2*, and *pap1 ogdh1/2*. Values are means ± SD of three biological replicates, and letters mark statistically significant differences (one-way ANOVA, $P < 0.01$).

domain and a conserved C-terminal domain) interacted with the JmjN domain–bearing N termini of the H3K27 demethylase family members, including ELF6, REF6, and JMJ13 (Fig. 4A and fig. S9A). Next, we conducted coimmunoprecipitations (co-IPs) using F_1 seedlings expressing functional OGDH1:HA and ELF6:Flag (*41*), and found that OGDH1:HA was immunoprecipitated by ELF6:Flag (Fig. 4B). Moreover, using epitope-tagged control proteins, we did not find any association of Flag with HA, ELF6 with HA, or OGDH1

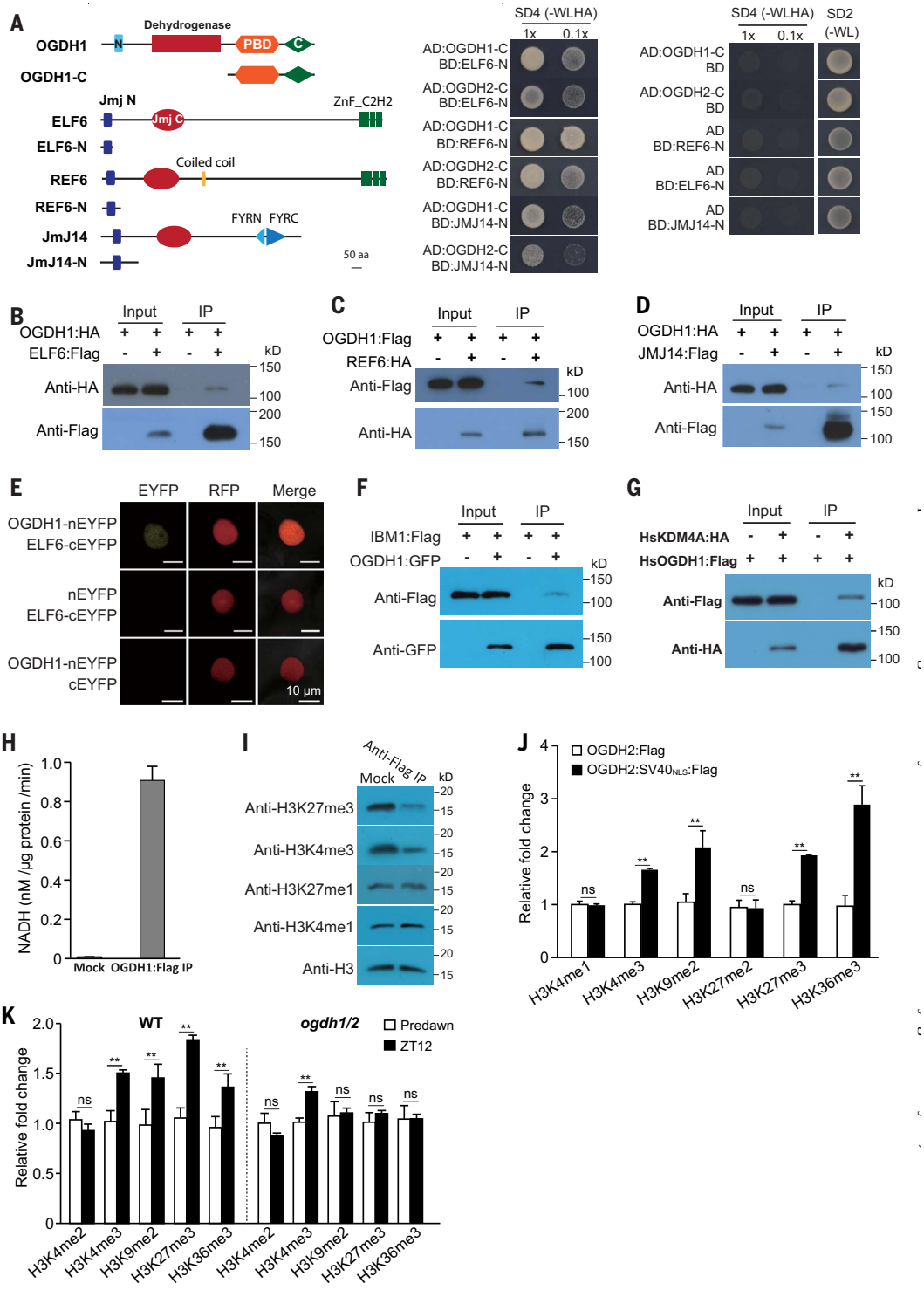


Fig. 4. OGDH subunits interact with JMJs and function to inhibit JMJ-mediated histone demethylations. (**A**) OGDH1 and OGDH2, through C terminus regions (OGDH-C), interacted with N terminus JmjN-bearing regions of ELF6, REF6, and JMJ14 in yeast cells. Left: schematic illustrations of proteins. N, conserved N-terminal motif in OGDH1; PBD, pyrimidine-binding domain; C, C-terminal domain. OGDH1-C (amino acids 601 to 1017) and OGDH2-C (amino acids 601 to 1025) were fused with GAL4-AD, and ELF6-N (amino acids 1 to 60), REF6-N (amino acids 1 to 90), and JMJ14-N (amino acids 1 to 182) were fused with GAL4-BD. Yeast cells were spread on the highly stringent selective medium of SD4 (−WLHA) without Trp (W), Leu (L), His (H), and adenine (A), and cells on SD2 (−WL) served as growth control. (**B** to **D**) Co-IP assays of OGDH1 with ELF6:Flag (B), REF6:HA (C), and JMJ14:Flag (D). Anti-Flag or anti-HA beads were used to immunoprecipitate protein complexes from F_1 seedlings expressing paired proteins, followed by immunoblotting. (**E**) BiFC analysis of OGDH1 interaction with ELF6 in *N. benthamiana* nuclei. Scale bars, 10 μm. (**F**) Co-IP of OGDH1:GFP with IBM1:Flag. OGDH1:GFP, complexed with IBM1:Flag, was immunoprecipitated by anti-GFP from F_1 seedlings expressing both proteins. (**G**) Co-IP of human OGDH1:Flag (HsOGDH1:Flag) with HsKDM4A:HA. Proteins were coexpressed transiently in human HEK 293F cells. (**H**) Oxidative decarboxylation of α-KG by the anti-Flag immunoprecipiated nuclear KGDH from *Arabidopsis* seedlings. Activity was calculated as nanomoles of NADH produced per microgram of protein per minute. Data are shown as means ± SD of three biological replicates. (**I**) Immunoblotting analysis of histone demethylation activities of JMJs coimmunoprecipiated with the nuclear KGDH. The precipitates by anti-Flag (recognizing OGDH1:Flag) were incubated with histones for demethylation assays. Data from one of two biological replicates with similar results are presented. (**J**) Changes in the levels of H3 methylated at K4, K9, K27, or K36 upon elevated KGDH nuclear targeting. Total histones extracted from *OGDH2:Flag* and *OGDH2:SV40$_{NLS}$:Flag* seedlings were analyzed by immunoblotting. (**K**) Relative levels of H3 methylated at K4, K9, K27, or K36 in WT and *ogdh1/2* seedlings at predawn and ZT12. Total histones were analyzed by immunoblotting. Values in (J) and (K) are means ± SD of three biological replicates analyzed for statistical significance by two-tailed Student's *t* test. $^{**}P < 0.01$; ns, not significant.

with Flag (fig. S9, F to H). These results confirm that OGDH1 indeed associates with the ELF6 H3K27 demethylase in seedlings. Furthermore, we performed bimolecular fluorescence complementation (BiFC) assays through transient coexpression of OGDH1-nEYFP and ELF6-cEYFP in tobacco epidermal cells, and found that OGDH1 physically associated with ELF6 in the nucleus, but not in the cytoplasm or mitochondria (Fig. 4E). Similarly, OGDH1 was found by co-IP to be in a complex with REF6 (Fig. 4C and fig. S9I), using F_1 seedlings expressing functional REF6:HA and OGDH1: Flag (fig. S9, B and C). These results together show that the

nuclear OGDH/KGDH is in a complex with several H3K27 demethylases.

The N-terminal JmjN domains exist widely in most JMJs in both plants and animals, and are involved in protein-protein interactions (*42*, *43*). We further found that the N-terminal JmjN domain–bearing region of the H3K4 demethylase JMJ14 (also known as PKDM7B) interacted with the C-terminal regions of OGDH1 and OGDH2 in yeast cells (Fig. 4A). Moreover, we conducted co-IPs using the F_1 seedlings expressing functional OGDH1:HA and JMJ14:Flag (*44*), and found that OGDH1 was immunoprecipitated by JMJ14:Flag. Thus, OGDH1 is indeed in a complex with JMJ14 in seedlings (Fig. 4D), consistent with OGDH/KGDH functioning to inhibit global H3K4me3 demethylation. Similarly, OGDH1 was found to be in a complex with the H3K9 demethylase IBM1 using yeast two-hybrid analysis and co-IPs with the F_1 seedlings expressing the functional IBM1:Flag (*45*) and OGDH1:GFP (Fig. 4F and fig. S9D), consistent with the function of OGDH/KGDH in inhibiting H3K9me2 demethylation.

Association between mammalian KGDH and JMJs

Most JMJs are evolutionarily conserved in the eukaryotes bearing histones with tails (*3*, *7*, *46*). In addition, the C-terminal regions of OGDH subunits are composed of a pyrimidine-binding domain and a C-terminal domain, which are highly conserved across eukaryotes (fig. S1A). Because a minor portion of KGDHs can be targeted into the nuclei of several types of mammalian cells (*21*, *47*, *48*), we investigated whether the association of nuclear OGDH/KGDH with JMJs may be conserved in mammals. Using yeast two-hybrid assays, we found that C-terminal regions of OGDH1 from human and mouse indeed interacted with the JmjN-bearing N termini of the human and mouse H3K9/H3K36 demethylase KDM4A, respectively (fig. S9E). Furthermore, we coexpressed the human OGDH1:Flag and KDM4A:HA in mammalian human embryonic kidney (HEK) 293F cells and found that these two proteins indeed associate with each other by co-IPs (Fig. 4G). There was no association of the human OGDH1 with HA or KDM4A with Flag (fig. S9, J and K); therefore, it is possible that the nuclear KDGH may interact and function together with JMJs to control histone demethylations in mammals in addition to plants.

JMJ-associated KGDH catalyzes oxidative decarboxylation of α-KG in the nucleus

To determine the biochemical function of nuclear KGDH, we immunoprecipitated KGDH from the nuclei isolated from the transgenic *Arabidopsis* seedlings expressing the functional OGDH1:Flag and further examined KGDH activity. We found that the immunoprecipitated KGDH complex from nuclei had potent oxidative decarboxylation activity (Fig. 4H). Next, we investigated whether active JMJs are co-IPed with the nuclear KGDH by measuring the demethylation activities of nuclear KGDH immunoprecipitations, and found that these precipitates demethylated H3K4me3 and H3K27me3 (substrates of JMJs), but not H3K27me1 or H3K4me1 (Fig. 4I). In addition, these activities were inhibited by succinate (fig. S10A), a known inhibitor of JMJ activity (*49*). These results reveal that active JMJs indeed associate with the KGDH holoenzyme in the nucleus.

Elevated KGDH in the nucleus inhibits histone demethylation by JMJs

To further confirm that nuclear KGDH acts to inhibit JMJ activities, we fused a potent nuclear localization signal from Simian Virus 40 (SV40) (*50*) at the C terminus of OGDH2 (OGDH2:NLS_{SV40}:Flag). The recombinant *OGDH2:NLS_{SV40}:Flag* was transformed into *Arabidopsis*. More OGDH2:NLS_{SV40}:Flag proteins were targeted to the nucleus compared with OGDH2:Flag (fig. S10B), which had minimal effects on the level of total cellular α-KG (fig. S10C). We found that elevated nuclear accumulation of OGDH2 in the *OGDH2:NLS_{SV40}:Flag* line resulted in increased levels of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 (preferential substrates of JMJs), but not of H3K4me1 or H3K27me2 (Fig. 4J). These results revealed that elevated nuclear accumulation OGDH2 (KGDH) inhibits histone demethylations by JMJs, reflecting that the nuclear KGDH holoenzyme, complexed with JMJs, catalyzes oxidative decarboxylation of α-KG to reduce its availability to JMJs.

Targeting KGDH into the nucleus is promoted by light, so we examined the levels of methylated H3 histones in early seedlings exposed to light for 12 hours [Zeitgeber time, 12 hours after lights on (ZT12)]. Compared with predawn, light exposure resulted in apparent increases, including in H3K4me3, H3K9me2, H3K27me3, and H3K36me3, among which the elevations of methyl H3K9, H3K27, and H3K36 were fully dependent on the activity of OGDH/KGDH (Fig. 4K). Thus, light promotes KGDH targeting into the nucleus to inhibit histone demethylations by JMJs in early seedlings.

In summary, the results thus far reveal that the nuclear KGDH holoenzyme, in complex with individual JMJs, catalyzes oxidative decarboxylation of α-KG to reduce its local availability to JMJs (on target chromatin), resulting in an inhibition of histone demethylations by JMJs in the *Arabidopsis* genome.

Nuclear KGDH occupies regions of target loci where histone methylations occur

We explored the genome-wide occupancy of OGDH1 by ChIP-seq using *OGDH1:Flag* seedlings. We found that 3528 loci were occupied by OGDH1:Flag across the *Arabidopsis* genome (Fig. 5, A to C, and fig. S11A). OGDH1 was enriched in gene-body regions of most target loci; in addition, it bound to 5′ transcribed regions proximal to the transcription start site (TSS) at part of the target loci (Fig. 5, A and B). Histone methylations, including H3K9me2, H3K27me3, and H3K36me3, are known to preferentially occur in gene-body regions in the *Arabidopsis* genome, whereas H3K4 trimethylation occurs in 5′ transcribed regions with a peak near the TSS (*39*). Thus, the spatial pattern of genome-wide OGDH1 occupancy is consistent with its function in inhibiting histone demethylations of H3K4me3, H3K9me2, H3K27me3, and H3K36me3 by JMJs. We further found that more than two-thirds of the OGDH1-bound loci (2454 of 3528) were marked with H3K4me3, and about one-fourth of these loci (915) were marked with H3K27me3 in *Arabidopsis* seedlings (Fig. 5C). The rest of OGDH1-bound loci may bear other lysine methylations such as H3K9me2 or H3K36me3. Together, these results reveal that nuclear OGDH/KGDH occupies 5′ transcribed and/or gene-body regions of target loci where histone methylations occur.

Next, we examined the changes of H3K27me3 at the 915 OGDH1-occupied loci (with H3K27me3 marks) in *ogdh1/2* relative to WT, which consist of 229 H3K27me3-bearing genes and 686 loci with both H3K4me3 and H3K27me3 (Fig. 5C). We found that upon loss of *OGDH1/2* function, most of the 229 H3K27me3-bearing loci occupied by OGDH1 displayed an overall reduction of H3K27me3 in gene-body regions where H3K27me3 was enriched, suggesting that OGDH/KGDH functions to inhibit H3K27 demethylation at these loci. At the 686 loci with H3K27me3 and H3K4me3, about half exhibited an H3K27me3 reduction in gene-body regions, whereas about one-third displayed a moderate increase in H3K27me3 (Fig. 5E). This indicates that at the loci with multiple methylation marks, OGDH/KGDH-mediated inhibition of JMJ-dependent histone demethylations may exert a complex effect on local chromatin features.

To further determine the role of *OGDH1* in transcriptional regulation of target genes, we compared the list of OGDH1-bound loci with the list of differentially expressed genes (DEGs) with a cutoff value of 1.8 in *ogdh1/2* (versus WT), and found that 174 of the 1008 up-regulated genes ($P = 4.8 \times 10^{-10}$, Fisher's exact test) and 263 of the 1572 down-regulated genes ($P = 5.5 \times 10^{-13}$) were bound by OGDH1 (Fig. 5D and table S4). Moreover, comparing the list of OGDH1-bound loci with DEGs with a cutoff value of 1.5 (in *ogdh1/2* versus WT) revealed that a total of 352 up-regulated genes and 402 down-regulated genes were bound by OGDH1 (fig. S11B and table S4). These results show that the chromatin-bound OGDH1 is required to promote or repress the expression of various genes in *Arabidopsis*.

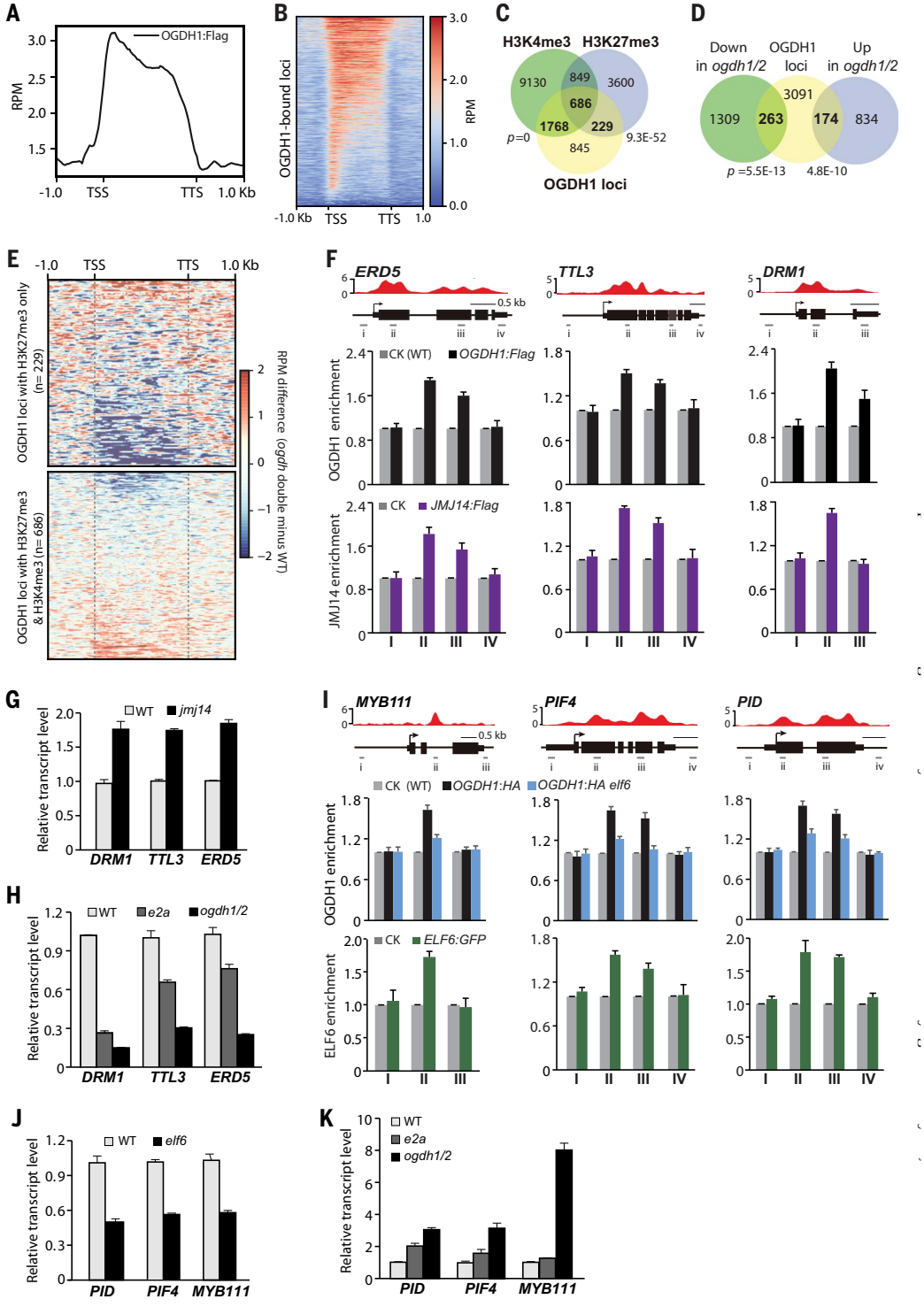


Fig. 5. OGDH/KGDH couples with JMJs to regulate histone demethylations and gene expression in the *Arabidopsis* genome. (**A** and **B**) Metagene plot (A) and heatmap (B) showing OGDH1:Flag occupancy at target loci in the *Arabidopsis* genome. (**C** and **D**) Overlapping of the OGDH1:Flag-bound loci with H3K4me3- and/or H3K27me3-modified loci (C) or the genes misregulated in *ogdh1/2* (D). Overlapping *P* values (Fisher's exact test) are indicated. (**E**) Heatmaps showing H3K27me3 changes at the OGDH1-bound loci with H3K27me3 (top) or H3K27me3 and H3K4me3 (bottom) in *ogdh1/2* (versus WT). Blue color indicates a reduction of H3K27me3 in *ogdh1/2* relative to WT (at a seedling stage). (**F**) ChIP-qPCR analysis of co-occupancy of OGDH1:Flag and JMJ14:Flag at *ERD5*, *TTL3*, and *DRM1*. Top: genome browser views of OGDH1:Flag occupancy (ChIP-seq data with RPM for read coverage) and schematic drawings of gene structures. Arrows indicate TSS and gray bars indicate ChIP-examined regions. Total chromatin was extracted from the seedings expressing the functional OGDH1:Flag and JMJ14:Flag. Relative fold enrichments of OGDH1:Flag or JMJ14:Flag over a background control (WT) are presented. (**G** and **H**) Relative transcript levels of *ERD5*, *TTL3*, and *DRM1* in the seedlings of WT, *jmj14* (G), *e2a-1*, and *ogdh1/2* (H). Relative levels to WT are presented. (**I**) Co-occupancy of OGDH1:HA and ELF6:GFP and dependence of OGDH1 enrichment on ELF6 at *MYB111*, *PIF4*, and *PID*, as revealed by ChIP-qPCR. Top: genome browser views of OGDH1:Flag occupancy at and schematic drawings of these three loci. Arrows indicate TSS and gray bars indicate ChIP-examined regions. Total chromatin was extracted from *OGDH1:HA*, *OGDH1:HA elf6*, and *ELF6:GFP* seedlings. Relative fold enrichments of OGDH1:HA or ELF6:GFP over a background control (WT) are presented. (**J** and **K**) Relative transcript levels of *MYB111*, *PIF4*, and *PID* in the seedlings of WT, *elf6* (J), *e2a-1*, and *ogdh1/2* (K). Relative levels to WT are presented. In (F) to (K), values are means ± SD of three biological replicates.

KGDH and JMJs regulate local histone demethylations and gene expression

To explore how nuclear KGDH functions together with JMJs at individual target loci to regulate their expression, we selected six OGDH1-bound loci modified with H3K4me3 or H3K27me3 for further examination. First, using ChIP coupled with quantitative polymerase chain reaction (ChIP-qPCR), we validated the occupancy of OGDH1:Flag in gene-body regions at three environment-responsive loci marked by H3K4me3, including *AT1G28330* (*DORMANCY-ASSOCIATED PROTEIN 1*; *DRM1*), *AT2G42580* (*TETRATRICOPETIDE-REPEAT THIOREDOXIN-LIKE 3*; *TTL3*), and *AT3G30775* (*EARLY RESPONSIVE TO DEHYDRATION 5*; *ERD5*) (Fig. 5F and table S3). Given the association of nuclear KGDH with the H3K4 demethylase JMJ14, we investigated whether JMJ14 may be coenriched at the OGDH1-bound regions. Indeed, we found that at *TTL3* and

ERD5, JMJ14 and OGDH1 co-occupied 5′ transcribed regions and the middle of gene bodies, whereas at *DRM1*, these proteins co-occupied a 5′ transcribed region (Fig. 5F).

We further found that upon loss of *OGDH1/2* function, levels of H3K4me3 in the regions co-occupied by OGDH1 and JMJ14 were decreased at *DRM1*, *TTL3*, and *ERD5* (fig. S11C), suggesting that OGDH (KGDH) acts to inhibit JMJ14-mediated H3K4 demethylation. In addition, we found that upon loss of *JMJ14* function, the expression of these three genes were all up-regulated (Fig. 5G), revealing that JMJ14-mediated demethylation of H3K4me3 (an active mark) represses their expression. Furthermore, we found that loss of KGDH function in either *ogdh1/2* or *e2a* resulted in repression of these three genes (Fig. 5H), revealing that elevated levels of local α-KG at these loci function to repress their expression. This is consistent with the reduced levels of H3K4me3 at these loci in *ogdh1/2* (fig. S11C). Together, these results reveal that KGDH reduces the local availability of α-KG to inhibit JMJ14-mediated H3K4 demethylation at *DRM1*, *TTL3*, and *ERD5* and thus promote their expression.

Next, we examined the co-regulation of KGDH with an H3K27 demethylase at three H3K27me3-marked loci, including *AT2G34650* (*PINOID*; *PID*), *PHYTOCHROME INTERACTING FACTOR 4* (*PIF4*; encoding a red light-responsive transcription factor; *ref 36*), and *MYB DOMAIN PROTEIN 111* (*MYB111*). *MYB111* is a key transcription factor that activates the expression of early structural genes in anthocyanin biosynthesis and functions together with the master regulator, *PAP1*, to activate anthocyanin biosynthesis in land plants (*51*). We first validated that OGDH1 was enriched in the gene-body regions of *PID*, *PIF4*, and *MYB111* (Fig. 5I). Given the physical association of OGDH1 with the H3K27 demethylase ELF6 in the nucleus, we further investigated with ChIP whether ELF6 may co-occupy with OGDH1 at *PID*, *PIF4*, and *MYB111* using the seedlings expressing a functional ELF6:GFP (fig. S11E). We found that ELF6 was indeed enriched in the gene-body regions where OGDH1 binds to at these loci (Fig. 5I). ELF6 is typically recruited to chromatin through its DNA-binding domains, association with a DNA-binding protein, and/or recognition of methyl histones (*41*, *52*). We reasoned that the KGDH holoenzyme may be recruited to *PID*, *PIF4*, and *MYB111* through the ELF6 protein bound to the chromatin regions at these loci. Indeed, we found that OGDH1 enrichment at these loci was dependent on ELF6 (Fig. 5I), revealing that ELF6 recruits OGDH/KGDH or stabilizes its association to the gene-body chromatin at these loci. This indicates that the enrichment of OGDH/KGDH at target chromatin regions depends on its coupled JMJs.

ELF6 catalyzes demethylation of H3K27me3, a repressive chromatin mark, to promote target gene expression. We found that *ELF6* was required for the expression of *PID*, *PIF4*, and *MYB111* (Fig. 5J). We further found that, in contrast to loss of *ELF6* function, loss of KGDH function in either *ogdh1/2* or *e2a* resulted in derepression of these three genes in seedlings (Fig. 5K), revealing that the elevated availability of local α-KG at these loci functions to promote their expression. Consistent with these findings, we found that the levels of H3K27me3 in the regions co-occupied by OGDH1 and ELF6 at *PID*, *PIF4*, and *MYB111* were decreased upon loss of *OGDH1/2* function (fig. S11D). Together, these results reveal that KGDH in the chromatin-bound ELF6-KGDH assembly consumes local α-KG to inhibit ELF6-mediated H3K27 demethylation and thus repress *PID*, *PIF4*, and *MYB111* expression.

ELF6 binds to 5′ transcribed and gene-body regions at *FLC* to mediate H3K27 demethylation, and thus promote *FLC* expression and inhibit the transition to flowering (*41*). Genome-wide analysis of OGDH1:Flag occupancy showed that it was enriched in the 5′ transcribed region and gene-body regions at *FLC* at the seedling stage, which was validated by ChIP-qPCR (fig. S11F). These results together suggest that KGDH, in association with ELF6, reduces the level of local α-KG at *FLC* to inhibit H3K27 demethylation by ELF6 and thus repress *FLC* expression and promote the floral transition in *Arabidopsis*.

In summary, chromatin-bound OGDH (KGDH) catalyzes α-KG decarboxylation, resulting in a reduction in its local availability to the KGDH-coupled JMJs and consequent inhibition of histone demethylations by JMJs, to regulate the expression of various target loci. Because different histone methylation marks are often read and interpreted by discrete effectors to repress or activate gene expression, the chromatin-bound KGDH can promote or repress target gene expression, depending on its coupled JMJs on target chromatin.

KGDH interacts with ELF6 to repress anthocyanin biosynthesis

To uncover the detailed molecular mechanism for *KGDH* regulation of *PAP1* expression and anthocyanin biosynthesis, we first conducted ChIP and found that both OGDH1:HA and OGDH2:GFP bound to *PAP1* chromatin in gene-body regions in seedlings (sampled at ZT6) (Fig. 6, A and B). *PAP1* chromatin bears the repressive H3K27me3 marks under normal growth conditions (table S2). We found that loss of *OGDH1/2* function resulted in an apparent reduction of H3K27me3 in the gene-body regions where both OGDH1 and OGDH2 bind (Fig. 6C). This suggests that the α-KG decarboxylase OGDH functions to inhibit H3K27 demethylation on *PAP1* chromatin.

Light promotes KGDH nuclear targeting. We found that a 6-hour light exposure led to increased OGDH1 binding to *PAP1* chromatin in young seedlings (ZT6 versus predawn) (fig. S12A). Moreover, we found that light exposure caused a strong increase of H3K27me3 specifically at the gene-body regions where OGDH1 binds, whereas only slight changes in H3K27me3 were observed in *ogdh1/2* (ZT6 versus predawn) (fig. S12B). This reveals that light-induced elevation in the enrichment of OGDH1 (and likely OGDH2) causes an inhibition of H3K27 demethylation at *PAP1*, reinforcing the notion that OGDH/KGDH consumes α-KG around target chromatin regions to inhibit H3K27 demethylation by JMJs at *PAP1*.

Next, we found that the H3K27 demethylase ELF6 (ELF6:GFP) was greatly enriched in the gene-body regions where both ODGH1 and OGDH2 bind (Fig. 6D). We further found that loss of *ELF6* function, in contrast to loss of *OGDH1/2* function, gave rise to strongly elevated levels of H3K27me3 in the gene-body regions bound by ELF6 and OGDH1/2 (Fig. 6C). Thus, the OGDH1/2-coupled ELF6 catalyzes H3K27me3 demethylation on *PAP1* chromatin. The loss of *ELF6* function led to moderately elevated levels of H3K27me3 in the regions upstream of *PAP1* gene body where ELF6 was not stably enriched (Fig. 6C), which may be attributed to a loss of transient ELF6 association or spreading of elevated H3K27me3 from the gene body.

We compared the levels of H3K27me3 at the gene-body regions of *PAP1* in WT, *ogdh1/2*, and *elf6 ogdh1 ogdh2* and found that the decrease of H3K27me3 in *ogdh1/2* was rescued by loss of *ELF6* function (Fig. 6C). Levels of H3K27me3 in the OGDH1/2–bound regions at *PAP1* were moderately lower in *elf6 ogdh1 ogdh2* than in *elf6*, suggesting that there may be another H3K27 demethylase functioning at *PAP1*. Indeed, we found that the ELF6 homolog JMJ13, a partner of the nuclear KGDH, was also enriched in *PAP1* gene-body regions, like ELF6 and OGDH1/2 (fig. S12C). This indicates that JMJ13 may partly mediate H3K27 demethylation at *PAP1*, particularly in *ogdh1/2* (with elevated α-KG).

We further determined the role of the KGDH-ELF6 assembly in *PAP1* regulation and anthocyanin biosynthesis. Consistent with the elevated levels of H3K27me3 at *PAP1* in *elf6*, *PAP1* expression was further repressed upon loss of *ELF6* function (Fig. 6E). Moreover, the great derepression of *PAP1* upon loss of *OGDH1/2* function was largely suppressed by loss of *ELF6* function in the *elf6 ogdh1 ogdh2* seedlings, and consequently anthocyanin accumulation was largely eliminated in this triple mutant (Fig. 6E and fig. S12D). These results reveal that KGDH-dependent *PAP1* repression mainly occurs through an inhibition of ELF6-mediated demethylation of H3K27me3 on gene-body chromatin at *PAP1*.

Thus far, the results have revealed that in response to light exposure, KGDH enters the

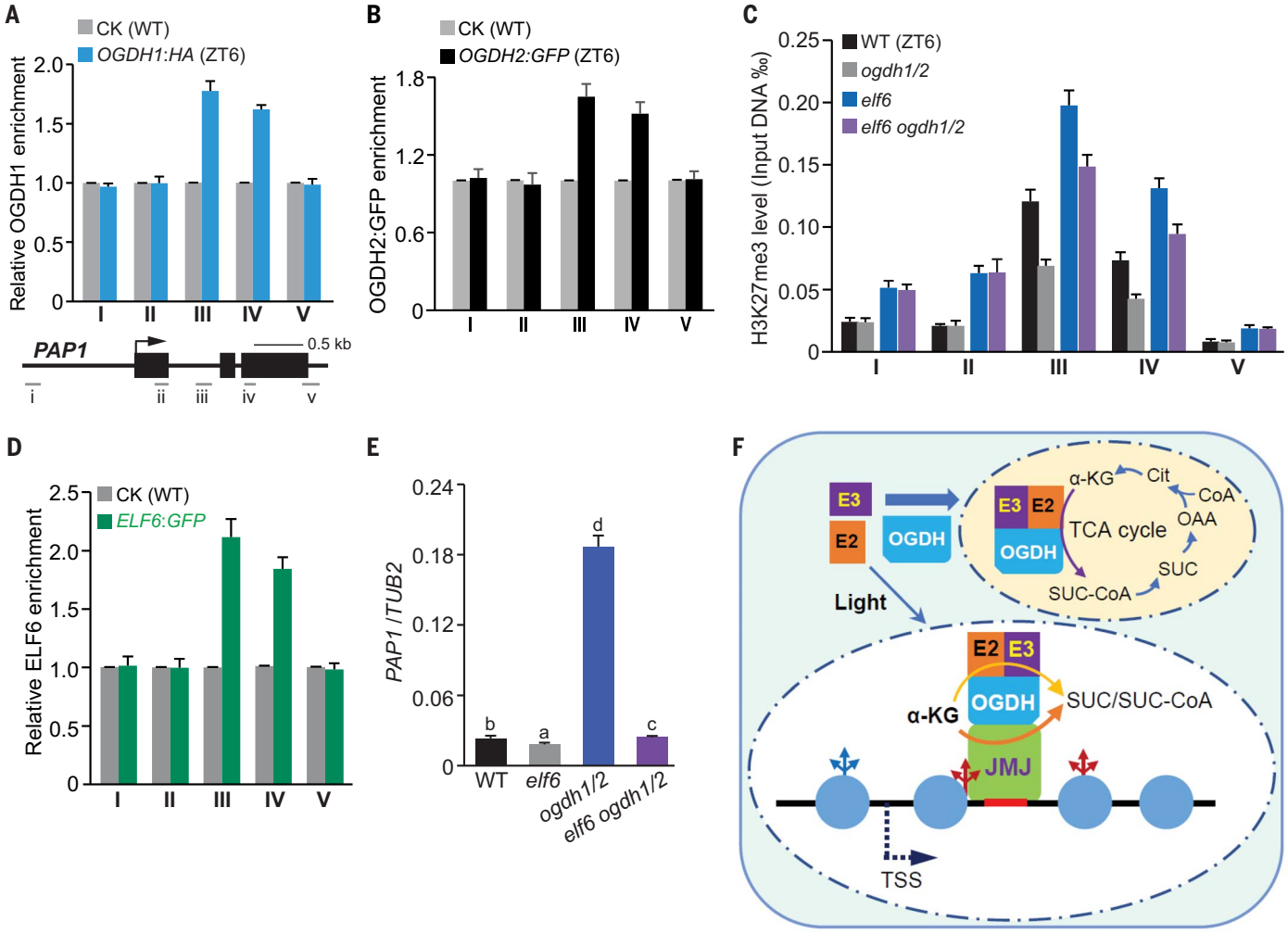


Fig. 6. KGDH couples with ELF6 to inhibit local H3K27 demethylation by ELF6 at *PAP1* to repress its expression. (**A** and **B**) ChIP-qPCR analysis of OGDH1:HA (A) and OGDH2:GFP (B) enrichment on *PAP1* chromatin. At bottom in (A) is a schematic drawing of the *PAP1* gene structure. Arrow indicates TSS and gray bars indicate ChIP-examined regions. Relative fold enrichments of OGDH1:HA or OGDH2:GFP over a background control (WT) at examined regions are presented. (**C**) ChIP analysis of H3K27me3 status on *PAP1* chromatin in the indicated seedlings. Levels of genomic fragments of each examined region (immunoprecipitated by anti-H3K27me3) were quantified and normalized to "input DNA." (**D**) ChIP analysis of ELF6:GFP enrichment on *PAP1* chromatin. Shown are relative fold enrichments of ELF6:GFP over a background control. (**E**) Transcript levels of *PAP1* in WT, *elf6*, *ogdh1/2*, and *elf6 ogdh1/2* seedlings. The transcripts were quantified by qPCR and normalized to *TUB2*. For (A) to (E), samples were harvested at ZT6, and values are means ± SD of three biological replicates. Letters in (E) mark statistically significant differences (one-way ANOVA, $P < 0.01$). (**F**) Working model for the dynamic control of histone demethylations by nuclear KGDH. Environmental signals such as light promote a minor portion of cytoplasmic KGDH entering the nucleus. KGDH forms an assembly with chromatin-bound JMJs, catalyzes α-KG decarboxylation, and thus limits its local availability to the JMJs coupled with KGDH, resulting in an inhibition of histone demethylations in target chromatin regions to regulate gene expression. The conversion of NAD^+ to NADH in α-KG decarboxylation may reduce the local NAD^+ pool, which likely has an effect on NAD^+-dependent histone modifications. In addition, succinyl-CoA is likely to be used as a substrate for histone succinylation. TCA cycle intermediates are indicated as follows: Cit, citrate; OAA, oxaloacetate; SUC, succinate; and SUC-CoA, succinyl-CoA.

nucleus and interacts with ELF6. The KGDH-ELF6 assembly is enriched at the gene-body chromatin of *PAP1*, where KGDH catalyzes α-KG decarboxylation. Reduction of local α-KG availability inhibits H3K27 demethylation by ELF6 on the gene-body chromatin at *PAP1*. This results in *PAP1* repression to prevent anthocyanin biosynthesis in *Arabidopsis* seedlings grown under proper conditions. Collectively, these findings illustrate a molecular mechanism underlying *PAP1* regulation by KGDH-ELF6, which exemplifies the KGDH-JMJ assembly–mediated regulation of histone demethylations and gene expression at target loci in the *Arabidopsis* genome (Fig. 6F).

Discussion

In this study, we have found that in *Arabidopsis* KGDH is targeted into the nucleus through the NLS in E2. Eukaryotic E2 subunits contain two highly conserved domains, a lipoyl attachment domain (LAD; near N terminal) and a lipoamide acyltransferase domain (LAT; C terminal) (*53*). The nuclear localization sequences located between LAD and LAT in the Arabdopsis E2a and E2b are not conserved in E2s from other eukaryotic organinsms such as green algae, moss, yeast, or fly, but these eukaryotic E2s bear a predicted NLS. This indicates that a small portion of the TCA cycle–associated holoenzyme KGDHs may be targeted into the nucleus in other eukaryotic organisms in addition to *Arabidopsis*. Indeed, it has been reported that a minor fraction of KGDH is localized in the nuclei of several types of mammalian cells (*21*, *47*, *48*). Given that the highly conserved C-terminal regions of OGDH subunits interact with JmjN-bearing N-terminal regions from JMJs in *Arabidopsis*, mouse, and human, it is

very likely that the association of nuclear KGDH with JMJs to control histone demethylation activities is evolutionarily conserved in eukaryotic organisms.

Nuclear KGDH targeting is regulated by environmental signals, and KGDH functions together with JMJs to regulate the expression of various environment-responsive genes in *Arabidopsis*. The H3K27 demethylases REF6 and JMJ13 regulate various environment-responsive genes in *Arabidopsis*, and the genes involved in responses to various stimuli are highly enriched in the REF6-bound loci (*15*, *52*, *54*). In addition, a large number of IBM1-regulated protein-coding genes are stress or environment responsive (*11*). Moreover, an analysis of *cis*-acting DNA elements located in the promoter regions of *JMJ*s in diverse plant species revealed that biotic and abiotic stress-related elements are rich in these regions, suggesting that the expression of certain *JMJ*s is regulated by environmental factors (*46*). These findings together reveal that JMJ-dependent control of the dynamics of histone methylations at environment-responsive loci may play a critical role in transcriptome reprogramming during environmental responses in plants.

In several human cancer cell lines, nuclear KGDH interacts with lysine acetyltransferase 2A (KAT2A) (*21*). The KGDH-coupled KAT2A succinylates histone 3 on lysine 79 (H3K79) in the proximal promoter regions of a subset of genes using the succinyl-coenzyme A generated locally by KGDH-dependent α-KG decarboxylation to promote gene expression in the cancer cells (*21*). In this study, we found that in *Arabidopsis*, the nuclear KGDH, in complex with various JMJs, regulates genome-wide gene expression primarily by inhibiting histone demethylations by JMJs. It will be interesting to explore whether the nuclear KGDH may couple with an *Arabidopsis* homolog of KAT2A to mediate H3K79 succinylation at a subset of loci, in addition to its general role in regulating histone demethylation activities of JMJs.

In summary, we have uncovered that the TCA cycle–associated enzyme KGDH enters the nucleus in response to environmental signals. Nuclear KGDH interacts with various JMJs, catalyzes oxidative decarboxylation of α-KG, and thus inhibits α-KG–dependent histone demethylations by JMJs to regulate genome-wide gene expression in plants. The association of nuclear KGDH with JMJs occurs in mammalian cells, suggesting that KGDH may function together with JMJs to control histone demethylation and gene expression in mammals in addition to plants.

Materials and Methods

Plant materials and growth condition

Arabidopsis mutants and transgenic lines used in this study are in the Columbia (Col-0) background. *elf6-3*, *ref6-1*, and *jmj14* were described previously (*8*, *55*). *ogdh1-1* (*Salk_072343*), *ogdh1-2* (*Salk_002473*), *e2a-1* (*Salk_134088*), and *e2b-1* (*Salk_005851*) were obtained from the *Arabidopsis* Biological Resource Center. *Arabidopsis* plants were grown in long days (LDs, 16 hours of light, 8 hours of dark) under cool white light with a relative humidity of ~60% at 22°C. *Nicotiana benthamiana* plants were grown in LDs at ~25°C.

Anthocyanin extraction and measurement

Anthocyanins were extracted and measured as described previously (*56*). Briefly, 11-day-old seedlings grown on half-strength Murashige and Skoog medium were harvested, followed by anthocyanin extraction with 1% HCl in methanol (v/v). Subsequently, 0.5 volume of deionized water and 1 volume of chloroform was added to remove chlorophyll. After centrifugation at 13,000*g* for 5 min, the supernatant was collected for spectrophotometric quantification at 530 and 657 nm. Relative content of anthocyanins was calculated by (A530 − A657)/g (fresh weight). Samples were analyzed in three biological replicates.

Metabolic profiling by liquid chromatography tandem mass spectrometry

Measuring of primary metabolites in seedlings by liquid chromatography–tandem mass spectrometry (LC-MS/MS) were conducted as previously described with minor modifications (*57*, *58*). Briefly, seedlings were harvested and ground into fine powder in liquid nitrogen, followed by metabolite extraction. For each sample, ~60 mg of powder was dissolved in 800 μl of extraction buffer (7 methanol: 3 chloroform, v/v, precooled at −20°C), and incubated at −20°C for 3 hours; subsequently, 560 μl of double-distilled water (ddH2O) was added, and 800 μl of supernatant was collected after centrifugation at 2200*g* for 10 min at 4°C, followed by another around of extraction with 800 μl of buffer (1 methanol: 1 ddH2O, v/v, precooled at 4°C). A total of 1.6 ml of supernatant per sample was filtered with a 0.2-μM nylon filter; subsequently, 1 ml was taken for MS/MS analysis, and 20 μl was used for quality control. The extractions were carried out on ice.

For the LC analysis, a Luna NH2 column [3 μm (particle), 100 mm (length) × 2 mm (inner diameter); Phenomenex)] was used. The LC gradient was set with eluent A (10 mM ammonium acetate and 5% v/v acetonitrile solution, adjusted to pH 9.5 with ammonia water) and eluent B (acetonitrile) as follows: 0 to 1 min, 15% A; 1 to 8 min, 15 to 70% A; 8 to 20 min, 70 to 95% A; 20 to 22 min, 95% A; and 22 to 25 min, 15% A. For the MS analysis, QTRAP 6500+ (AB Sciex) was used in the MRM model, and all parameters were derived from previous studies with optimization (*57*, *58*). The concentration of each metabolite was calculated using a "concentration-peak area" curve of standard samples and presented as nanomoles per gram fresh weight.

Plasmid construction

To construct *OGDH1-GUS* and *OGDH2-GUS*, the genomic fragment of *OGDH1* (the 2-kb promoter plus the 1.8-kb coding sequence) and *OGDH2* (the 1.4-kb promoter plus the 1.9-kb coding sequence) were fused upstream with the coding region for *GUS* in *pMDC162* (*59*), using Gateway technology (Invitrogen). To create *pOGDH1-OGDH1: HA*, a genomic *OGDH1* fragment consisting of 2.0-kb promoter plus the 3.8-kb entire genomic coding sequence (without the stop codon) was first fused with a 3x HA tag sequence, which was subsequently cloned into *pHGW* (*60*) using Gateway technology. To create *pOGDH1-OGDH1:GFP*, the 5.8-kb genomic *OGDH1* fragment was fused upstream with the coding region for *GFP* in *pMDC110* (*59*) using Gateway technology. For *pOGDH2-OGDH2:GFP* construction, a genomic *OGDH2* fragment consisting of 1.4-kb promoter plus the 4.3-kb entire genomic coding sequence (without the stop codon) was fused upstream with *GFP* in *pMDC110*.

To construct *p35S-OGDH1:Flag* and *p35S-OGDH2:Flag*, the full-length genomic coding sequences for *OGDH1* (3.8 kb) and *OGDH2* (4.3-kb) without the stop codon were first fused with a 3x *Flag*, which were subsequently cloned downstream of the 35S promoter in *pMDC32* (*59*). To create *p35S-OGDH2:SV40$_{NLS}$-Flag*, the full-length OGDH2 genomic coding sequence (4.3 kb, without the stop codon) was first fused with the nuclear location signal of Pro-Lys-Lys-Lys-Arg-Lys-Val from SV40 (*50*); subsequently, a 3x *Flag* tag sequence was added to the fusion (at 3′ end), which was then cloned downstream of the 35S promoter in *pMDC32*.

To create *pE2a-E2a:GFP*, a genomic fragment of E2a consisting of 1.6-kb promoter plus 3.2-kb genomic coding sequence (without stop codon) was fused with *GFP* in *pMDC110*. For *pE2a-mtE2a-1:GFP* and *pE2a-mtE2a-2:GFP* constructions, the 4.8 genomic fragment of *E2a* bearing mutations for E2a-1 (R236A, R237A, R242A and R244A) or mutations for E2a-2 (K221A, S223A, K232A and R234A), was fused upstream with *GFP* in *pMDC110*. To construct *p35S-E3a:HA*, the full-length E3 genomic sequence (2.1 kb, without the stop codon) was first fused with a 3xHA tag, and the fusion was subsequently cloned downstream of the 35S promoter in *pMDC32*.

To construct *pREF6-REF6:HA*, a genomic fragment of *REF6* consisting of 1.7-kb promoter plus the 5.2-kb entire genomic coding sequence (without the stop codon) was first fused with a 3xHA tag, and then the fusion was cloned into *pHGW* (*60*). For *pELF6-ELF6:GFP* construction, a genomic fragment of *ELF6* consisting of the 0.7-kb promoter

plus the entire 5.5-kb genomic coding sequence (without the stop codon) was fused upstream with *GFP* in *pMDC110*. To create *pJMJ13-JMJ13:Flag*, a genomic fragment of *JMJ13* consisting of the 2.1-kb promoter plus the entire 5.5-kb genomic coding sequence (without the stop codon) was first fused with a 3x *Flag* tag sequence [as described in (*15*)], and then the fusion was cloned into *pHGW*. For *p35S-HTA9:RFP* construction, the full-length genomic sequence of *HTA9* (encoding H2A.Z; 0.83 kb) without the stop codon was cloned into a modified *pCAMBIA2300* vector, resulting in an *HTA9* fusion with *RFP*.

To construct *Hs_OGDH1:Flag* and *Hs_KDM4A:HA*, the coding regions of *Hs_OGDH1* and *Hs_KDM4A* were fused with 3x *Flag* and 3x *HA*, respectively, and the fusions were subsequently cloned into *pcDNA3.1* (GenScript Biotech) for transient expression in mammalian cells. For *p35S-GST:Flag* and *p35S-GFP:HA* constructions, *GST:3xFlag* and *GFP:3xHA* were cloned downstream of the *35S* promoter in *pB2GW7* (*60*).

Cellular protein fractionation

Cellular fractionation assays were conducted as described previously (*61*). Briefly, seedlings were ground into fine powder in liquid nitrogen and lysed in a fractionation lysis buffer [20 mM Tris-HCl, pH 7.5, 2 mM EDTA, 2.5 mM $MgCl_2$, 20 mM KCl, 5 mM dithiothreitol (DTT), 25% glycerol, 250 mM sucrose, and 1× Roche protease inhibitor cocktail], followed by centrifugation at 1500*g* for 10 min at 4°C. Subsequently, cytoplasmic proteins from the supernatant and nuclear proteins from the pellet were extracted, followed by immunoblotting. The cytoplasmic protein marker Actin and nuclear protein marker histone 3 (H3) were detected with anti-Actin (Agrisera, AS132640) and anti-H3 (Millipore, 07-690), respectively. Nuclear fractions were mitochondrion free (fig. S6C).

Yeast two-hybrid assay

Yeast two-hybrid assays to examine protein-protein interactions were conducted using the Matchmaker Gold System (Clontech). Coding sequences of genes of interest were cloned into *pGADT7* or *pGBKT7*. Paired plasmids were introduced into the yeast strain *AH109*. Subsequently, the transformed yeast cells were spread on solid dropout medium without leucine/tryptophan/adenine/histidine (–L/W/A/H) to determine protein-protein interactions.

BiFC assay

The full-length coding sequences for OGDH1 and ELF6 were fused in frame with an N-terminal and a C-terminal EYFP fragment, respectively; subsequently, this pair of plasmids together with the nuclear marker plasmid of *HTA9:RFP* were introduced into *N. benthamiana* epidermal cells by *Agrobacterium*-mediated transient transformation as described previously (*62*). Within 48 hours after inoculation, the fluorescence of EYFP and RFP was imaged with a Leica TCS SP8 SMD microscope.

Co-IP

Co-IP experiments with *Arabidopsis* seedlings or *N. benthamiana* leaves expressing proteins of interest were conducted as previously described with minor modifications (*63*). Briefly, *Arabidopsis* seedlings were first treated with 10 μM MG132 (Sigma, M8699-1MG) for 3 hours, and subsequently harvested for total protein extraction. The extracts were immunoprecipitated with anti-HA beads (Sigma, SAE0197), anti-Flag beads (Sigma, M8823), or anti-GFP beads (Abcam, ab193983), followed by Western blotting to examine the presence or abundance of the proteins of interest in the precipitates.

HEK 293F cells (*64*) were cultured and transfected using polyethylenimine as described previously (*65*). Transfected cells were lysed in the NETN buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl, pH 8.0, and 0.5% Nonidet P-40), followed by total protein extraction as described previously (*66*). Subsequently, the extracts were immunoprecipitated with anti-HA beads (Sigma, SAE0197), followed by Western blotting.

Histone extraction and immunoblotting

Total histones were extracted and analyzed as previously described (*67*). Briefly, histones were extracted from 11-day-old seedlings, followed by immunoblotting with anti-H3K4me1 (Abcam, ab8895), anti-H3K4me2 (Millipore, 07-030), anti-H3K4me3 (Millipore, 05-745), anti-H3K9me2 (Millipore, D5567), anti-H3K27me2 (Millipore, 07-452), anti-H3K27me3 (Millipore, 07-449), anti-H3K36me3 (Millipore, ABE435), and anti-H3 (Millipore, 07-690). Blotting signals or band intensities were quantified using ImageJ software.

RNA analysis by reverse transcription coupled with qPCR

Total RNAs were extracted from about 11-d-old seedlings using RNeasy Plus mini kit (Qiagen), following the manufacturer's instructions, and reverse transcription (RT)–qPCR assays were performed as previously described (*67*). Briefly, 2 μg of total RNA was reversely transcribed to synthesize cDNAs, followed by real-time qPCR on an ABI QuantStudio real-time PCR system using a master mix of SYBR Green (Bio-Rad). The constitutively expressed *TUB2* served as an internal control. Three biological replicates were conducted for each sample, and primers are listed in table S5.

RNA-seq analysis

RNA-seq analysis was conducted as previously described (*67*). Briefly, total RNAs were extracted from seedlings using the RNeasy Plus mini kit (Qiagen), followed by cDNA library construction with the NEBNext Ultra Directional RNA Library Prep kit for Illumina (New England BioLabs, E7420L). cDNA libraries were sequenced on an Illumina NovaSeq6000 sequencer. Three biological replicates for each sample were conducted. RNA-seq data were processed as previously described (*67*), and DEGs were identified with the edgeR package (FDR < 0.05) (*68*).

Immunoprecipitation of nuclear OGDH/KGDH complex

About 1 g of seedlings expressing OGDH1:Flag were harvested, followed by nuclear isolation as previously described (*69*) with minor modifications. Briefly, frozen seedlings were ground into fine powder and dissolved in ice-cold nuclei isolation buffer [20 mM HEPES, pH 8, 250 mM sucrose, 1 mM $MgCl_2$, 5 mM KCl, 40% glycerol, 0.25% Triton X-100, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and 0.1% 2-mercaptoethanol], followed by filtering through Miracloth and centrifugation to separate nuclei from tissue debris. After several washes with ice-cold nuclei isolation buffer, intact nuclei were collected, followed by total nuclear protein isolation using an extraction buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.2% Triton X-100, 0.2% Nonidet P-40, 1 mM PMSF, and 1× Roche protease inhibitor cocktail). Nuclear KGDH complexes were immunoprecipiated with anti-Flag beads (Sigma, M8823), followed by several washes with a washing buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 8.0% glycerol, and 1 mM DTT).

Measurement of KGDH activity

The enzymatic activity of nuclear KGDH was measured using the α-Ketoglutarate Dehydrogenase Activity Colorimetric Assay Kit (Sigma, MAK189) according to the manufacturer's instructions with minor modifications. Briefly, the anti-Flag beads carrying OGDH1:Flag/KGDH and mock (precipitates from WT by anti-Flag beads) were incubated with the KGDH assay buffer, KGDH developer, and KGDH substrate from the kit. The absorbance of A450 was measured with a Varioskan Flash microplate reader (Thermo Scientific) at different time points of reaction at 37°C. The amount of NADH (nanomoles/well) generated by the KGDH holoenzyme between time point 1 and time point 2 was calculated according to a standard curve. Total proteins per reaction were measured with a Bradford protein assay kit (Sangon Biotech) according to the manufacturer's instructions. KGDH activity was calculated as nanomoles of NADH produced per microgram of protein per minute.

To measure KGDH activity in total nuclear proteins, nuclei were first extracted from about 1 g of seedlings of WT, *e2a-1*, and *ogdh1/2* as described previously (*69*). Subsequently, the

nuclei resuspended in the KGDH assay buffer sample were sonicated, followed by centrifugation at 10,000*g* for 5 min. Total nuclear proteins in the supernatants were used for α-KG decarboxylation and KGDH activity assays.

Histone demethylation assay

The histone demethylation activities of the nuclear OGDH/KGDH–associated JMJs were measured as previously described (*55*). Briefly, the anti-Flag beads carrying OGDH1-Flag/KGDH and mock were incubated with calf thymus histones (Sigma) in an assay buffer [20 mM Tris/HCl pH 7.5, 150 mM NaCl, 50 μM $(NH_4)_2Fe(SO_4)_2$, 1 mM α-ketoglutarate, and 2 mM ascorbic acid] for 4.5 hours at 37°C. In separate reactions, succinate (10 mM) was incubated briefly with the OGDH1-Flag immunoprecipitates before adding other reaction mixtures (*49*). The histones recovered from the reactions were separated by SDS–polyacrylamide gel electrophoresis and transferred to 0.2-μm nitrocellulose membranes, followed by immunoblotting with anti-trimethyl H3K4 (Millipore, 05-745), anti-monomethyl H3K4 (abcam, ab8895), anti-trimethyl H3K27 and anti-monomethyl H3K27 (Millipore, 07-448), or anti-H3 (Millipore, 07-690).

ChIP-seq

ChIP-seq experiments to analyze genome-wide occupancy of OGDH1:Flag were performed as described previously with minor modifications (*70*). Briefly, 11-day-old seedlings expressing OGDH1:Flag (pretreated with 10 μM MG132) were fixed with 1% formaldehyde in a fixation buffer (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, and 1 mM PMSF), followed by nucleus isolation as described previously (*69*). Subsequently, nuclei were resuspended in a lysis buffer, followed by chromatin extraction and sonication. OGDH1:Flag-bound chromatin was immunoprecipitated with anti-Flag beads (Sigma, M8823). After de-cross-linking, DNA fragments were purified with the QIAquick PCR purification kit. DNA libraries for sequencing were constructed using an NEBNext UltraTM II DNA Library Prep Kit for Illumina (New England BioLabs), and sequenced on an Illumina NovaSeq6000 platform. Two biologically independent samples were analyzed.

ChIP-seq analysis of H3K27me3 in WT and *ogdh1 ogdh2* was conducted as described previously (*70*). Briefly, WT and *ogdh1 ogdh2* seedlings were fixed, followed by nucleus isolation. Total chromatin was extracted and sonicated, followed by immunoprecipitations with anti-H3K27me3 (Millipore, 07-449). After de-cross-linking, DNA fragments were purified; subsequently, DNA libraries for sequencing were constructed and sequenced on an Illumina NovaSeq6000 platform. Three biologically independent samples were analyzed for each genotype.

Processing of ChIP-seq data

Raw sequencing data of OGDH1:Flag ChIP were cleaned with trim_galore (version 0.4.4) (https://github.com/FelixKrueger/TrimGalore), and adapters and reads with quality lower than 30 and length shorter than 50 bp were filtered. Reads were mapped to the TAIR10 reference genome using bowtie2 (version 2.2.9) (*71*) with default parameters, and the mapping files were filtered by samtools (version: 0.1.19) (*72*). Peaks were called using MACS2 (parameters:–gsize 1.19e8–qvalue 0.01) (*73*). The coverage calculation of alignments was performed using bamCoverage from deepTools2. To profile peak distribution, computeMatrix (parameters: -a 1000 -b 1000 -m 2000) and plotProfile (parameters:–plotHeight 8–plotWidth 12 –perGroup) from deepTools were used. To identify the loci where the peaks are located, ChIPSeeker (*74*) was used to annotate the peaks, and the genes with peaks were visualized by the WashU Epigenome Browser (*75*). In addition, the heatmaps of peaks were plotted using plotHeatmap from deepTools. To examine the consistency between replicates, the average scores of bam files were obtained using multiBigwigSummary from deepTools2 (version version 3.5.1) (*76*), and correlation analysis was performed with plotCorrelation from deepTools (parameters:–corMethod pearson—skipZeros).

ChIP-seq data of H3K27me3 were processed and analyzed as previously described (*70*). Enriched peaks were called by MACS with the cut-off *P* value of 0.05, and read coverage per 1× sequencing depth (reads per genomic content) or reads per million (at single-nucleotide resolution) over target loci were scored. The reads per million score differences of H3K27me3 enrichment between *ogdh1/2* and WT at the examined loci were calculated with computeMatrix from deepTools and visualized using plotHeatmap from deepTools.

ChIP-qPCR

ChIP assays were performed as described previously (*70*) with minor modifications. Briefly, the seedlings expressing OGDH1:HA, OGDH1:Flag, OGDH2:GFP, JMJ14:Flag, or ELF6:GFP were fixed with 1% formaldehyde in the fixation buffer, followed by nucleus isolation as described previously (*69*). Total chromatin was extracted and sonicated, followed by immunoprecipitations with anti-HA (Sigma, H9658), anti-Flag beads (Sigma, M8823), or anti-GFP (Abcam, ab290). After de-cross-linking, DNA fragments was purified with the QIAquick PCR purification kit. Subsequently, DNA fragments of interest were quantified on an ABI QuantStudio Real-Time PCR System using a master mix of SYBR Green (Bio-Rad). Three biologically independent samples were analyzed in each experiment, and each sample was quantified in triplicate. Primers are listed in table S5. Relative enrichment of a protein of interest at each examined region was first normalized to the endogenous control *TUB2*, followed by normalization to a background control (WT).

REFERENCES AND NOTES

1. A. Jambhekar, A. Dhall, Y. Shi, Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019). doi: 10.1038/s41580-019-0151-1; pmid: 31267065
2. J. C. Black, C. Van Rechem, J. R. Whetstine, Histone lysine methylation dynamics: Establishment, regulation, and biological impact. *Mol. Cell* **48**, 491–507 (2012). doi: 10.1016/j.molcel.2012.11.006; pmid: 23200123
3. S. M. Kooistra, K. Helin, Molecular mechanisms and potential functions of histone demethylases. *Nat. Rev. Mol. Cell Biol.* **13**, 297–311 (2012). doi: 10.1038/nrm3327; pmid: 22473470
4. S. Markolovic *et al*., Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr. Opin. Struct. Biol.* **41**, 62–72 (2016). doi: 10.1016/j.sbi.2016.05.013; pmid: 27309310
5. R. J. Klose, Y. Zhang, Regulation of histone methylation by demethylimination and demethylation. *Nat. Rev. Mol. Cell Biol.* **8**, 307–318 (2007). doi: 10.1038/nrm2143; pmid: 17342184
6. X. Chen, Y. Hu, D. X. Zhou, Epigenetic gene regulation by plant Jumonji group of histone demethylase. *Biochim. Biophys. Acta* **1809**, 421–426 (2011). doi: 10.1016/j.bbagrm.2011.03.004; pmid: 21419882
7. F. Lu *et al*., Comparative analysis of JmjC domain-containing proteins reveals the potential histone demethylases in *Arabidopsis* and rice. *J. Integr. Plant Biol.* **50**, 886–896 (2008). doi: 10.1111/j.1744-7909.2008.00692.x; pmid: 18713399
8. B. Noh *et al*., Divergent roles of a pair of homologous jumonji/zinc-finger-class transcription factor proteins in the regulation of *Arabidopsis* flowering time. *Plant Cell* **16**, 2601–2613 (2004). doi: 10.1105/tpc.104.025353; pmid: 15377760
9. Z. Yang *et al*., Structure of the *Arabidopsis* JMJ14-H3K4me3 complex provides insight into the substrate specificity of KDM5 subfamily histone demethylases. *Plant Cell* **30**, 167–177 (2018). doi: 10.1105/tpc.17.00666; pmid: 29233856
10. D. Jiang, W. Yang, Y. He, R. M. Amasino, *Arabidopsis* relatives of the human lysine-specific Demethylase1 repress the expression of *FWA* and *FLOWERING LOCUS C* and thus promote the floral transition. *Plant Cell* **19**, 2975–2987 (2007). doi: 10.1105/tpc.107.052373; pmid: 17921315
11. S. Inagaki *et al*., Gene-body chromatin modification dynamics mediate epigenome differentiation in *Arabidopsis*. *EMBO J.* **36**, 970–980 (2017). doi: 10.15252/embj.201694983; pmid: 28100676
12. J. Cheng *et al*., H3K9 demethylases IBM1 and JMJ27 are required for male meiosis in *Arabidopsis thaliana*. *New Phytol.* **235**, 2252–2269 (2022). doi: 10.1111/nph.18286; pmid: 35638341
13. F. Lu, X. Cui, S. Zhang, T. Jenuwein, X. Cao, *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat. Genet.* **43**, 715–719 (2011). doi: 10.1038/ng.854; pmid: 21642989
14. W. Yan *et al*., Dynamic and spatial restriction of Polycomb activity by plant histone demethylases. *Nat. Plants* **4**, 681–689 (2018). doi: 10.1038/s41477-018-0219-5; pmid: 30104650
15. S. Zheng *et al*., The *Arabidopsis* H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. *Nat. Commun.* **10**, 1303 (2019). doi: 10.1038/s41467-019-09310-x; pmid: 30899015
16. Y. Yan *et al*., A MYB-domain protein EFM mediates flowering responses to environmental cues in *Arabidopsis*. *Dev. Cell* **30**, 437–448 (2014). doi: 10.1016/j.devcel.2014.07.004; pmid: 25132385
17. R. H. Bloomer *et al*., The *Arabidopsis* epigenetic regulator ICU11 as an accessory protein of Polycomb Repressive Complex 2. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16660–16666 (2020). doi: 10.1073/pnas.1920621117; pmid: 32601198
18. A. Salminen, A. Kauppinen, M. Hiltunen, K. Kaarniranta, Krebs cycle intermediates regulate DNA and histone methylation: Epigenetic impact on the aging process. *Ageing Res. Rev.* **16**, 45–65 (2014). doi: 10.1016/j.arr.2014.05.004; pmid: 24910305
19. H. Abla, M. Sollazzo, G. Gasparre, L. Iommarini, A. M. Porcelli, The multifaceted contribution of α-ketoglutarate to tumor progression: An opportunity to exploit? *Semin. Cell Dev. Biol.* **98**, 26–33 (2020). doi: 10.1016/j.semcdb.2019.05.031; pmid: 31175937

20. A. Nunes-Nesi, W. L. Araújo, T. Obata, A. R. Fernie, Regulation of the mitochondrial tricarboxylic acid cycle. *Curr. Opin. Plant Biol.* **16**, 335–343 (2013). doi: 10.1016/j.pbi.2013.01.004; pmid: 23462640
21. Y. Wang *et al*., KAT2A coupled with the α-KGDH complex acts as a histone H3 succinyltransferase. *Nature* **552**, 273–277 (2017). doi: 10.1038/nature25003; pmid: 29211711
22. J. A. Condori-Apfata *et al*., The Arabidopsis E_1 subunit of the 2-oxoglutarate dehydrogenase complex modulates plant growth and seed production. *Plant Mol. Biol.* **101**, 183–202 (2019). doi: 10.1007/s11103-019-00900-3; pmid: 31286324
23. J. M. Schvartzman, C. B. Thompson, L. W. S. Finley, Metabolic regulation of chromatin modifications and gene expression. *J. Cell Biol.* **217**, 2247–2259 (2018). doi: 10.1083/jcb.201803061; pmid: 29760106
24. X. Cui, Y. Zheng, Y. Lu, E. Issakidis-Bourguet, D. X. Zhou, Metabolic control of histone demethylase activity involved in plant response to high temperature. *Plant Physiol.* **185**, 1813–1828 (2021). doi: 10.1093/plphys/kiab020; pmid: 33793949
25. B. W. Carey, L. W. Finley, J. R. Cross, C. D. Allis, C. B. Thompson, Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells. *Nature* **518**, 413–416 (2015). doi: 10.1038/nature13981; pmid: 25487152
26. W. Xu *et al*., Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011). doi: 10.1016/j.ccr.2010.12.014; pmid: 21251613
27. T. Laukka, M. Myllykoski, R. E. Looper, P. Koivunen, Cancer-associated 2-oxoglutarate analogues modify histone methylation by inhibiting histone lysine demethylases. *J. Mol. Biol.* **430** (18 Pt B), 3081–3092 (2018). doi: 10.1016/j.jmb.2018.06.048; pmid: 29981745
28. N. Lamadema, S. Burr, A. C. Brewer, Dynamic regulation of epigenetic demethylation by oxygen availability and cellular redox. *Free Radic. Biol. Med.* **131**, 282–298 (2019). doi: 10.1016/j.freeradbiomed.2018.12.009; pmid: 30572012
29. J. F. Couture, E. Collazo, P. A. Ortiz-Tello, J. S. Brunzelle, R. C. Trievel, Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695 (2007). doi: 10.1038/nsmb1273; pmid: 17589523
30. J. A. Condori-Apfata *et al*., Downregulation of the E2 subunit of 2-oxoglutarate dehydrogenase modulates plant growth by impacting carbon-nitrogen metabolism in *Arabidopsis thaliana*. *Plant Cell Physiol.* **62**, 798–814 (2021). doi: 10.1093/pcp/pcab036; pmid: 33693904
31. L. Chen, B. Hu, Y. Qin, G. Hu, J. Zhao, Advance of the negative regulation of anthocyanin biosynthesis by MYB transcription factors. *Plant Physiol. Biochem.* **136**, 178–187 (2019). doi: 10.1016/j.plaphy.2019.01.024; pmid: 30685697
32. A. M. LaFountain, Y. W. Yuan, Repressors of anthocyanin biosynthesis. *New Phytol.* **231**, 933–949 (2021). doi: 10.1111/nph.17397; pmid: 33864686
33. C. T. Armstrong, J. L. Anderson, R. M. Denton, Studies on the regulation of the human E1 subunit of the 2-oxoglutarate dehydrogenase complex, including the identification of a novel calcium-binding site. *Biochem. J.* **459**, 369–381 (2014). doi: 10.1042/BJ20131664; pmid: 24495017
34. R. Nagaraj *et al*., Nuclear localization of mitochondrial TCA cycle enzymes as a critical step in mammalian zygotic genome activation. *Cell* **168**, 210–223.e11 (2017). doi: 10.1016/j.cell.2016.12.026; pmid: 28086092
35. A. N. Nguyen Ba, A. Pogoutse, N. Provart, A. M. Moses, NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10**, 202 (2009). doi: 10.1186/1471-2105-10-202; pmid: 19563654
36. M. C. Cheng, P. K. Kathare, I. Paik, E. Huq, Phytochrome signaling networks. *Annu. Rev. Plant Biol.* **72**, 217–244 (2021). doi: 10.1146/annurev-arplant-080620-024221; pmid: 33756095
37. J. Xiao *et al*., *Cis* and *trans* determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. *Nat. Genet.* **49**, 1546–1552 (2017). doi: 10.1038/ng.3937; pmid: 28825728
38. Y. Zhou *et al*., Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in *Arabidopsis*. *Nat. Genet.* **50**, 638–644 (2018). doi: 10.1038/s41588-018-0109-9; pmid: 29700471
39. F. Roudier *et al*., Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* **30**, 1928–1938 (2011). doi: 10.1038/emboj.2011.103; pmid: 21487388
40. S. D. Michaels, R. M. Amasino, *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999). doi: 10.1105/tpc.11.5.949; pmid: 10330478
41. Z. Li, Y. Ou, Z. Zhang, J. Li, Y. He, Brassinosteroid signaling recruits histone 3 lysine-27 demethylation activity to *FLOWERING LOCUS C* chromatin to inhibit the floral transition in *Arabidopsis*. *Mol. Plant* **11**, 1135–1146 (2018). doi: 10.1016/j.molp.2018.06.007; pmid: 29969683
42. S. Tronnersjö *et al*., The jmjN and jmjC domains of the yeast zinc finger protein Gis1 interact with 19 proteins involved in transcription, sumoylation and DNA repair. *Mol. Genet. Genomics* **277**, 57–70 (2007). doi: 10.1007/s00438-006-0171-3; pmid: 17043893
43. M. Levin, M. Stark, Y. G. Assaraf, The JmjN domain as a dimerization interface and a targeted inhibitor of KDM4 demethylase activity. *Oncotarget* **9**, 16861–16882 (2018). doi: 10.18632/oncotarget.24717; pmid: 29682190
44. Y. Q. Ning *et al*., Two novel NAC transcription factors regulate gene expression and flowering time by associating with the histone demethylase JMJ14. *Nucleic Acids Res.* **43**, 1469–1484 (2015). doi: 10.1093/nar/gku1382; pmid: 25578968
45. M. Lei *et al*., *Arabidopsis* EDM2 promotes IBM1 distal polyadenylation and regulates genome DNA methylation patterns. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 527–532 (2014). doi: 10.1073/pnas.1320106110; pmid: 24248388
46. S. Ma *et al*., Evolutionary history and functional diversification of the JmjC domain-containing histone demethylase gene family in plants. *Plants* **11**, 1041 (2022). doi: 10.3390/plants11081041; pmid: 35448769
47. E. Kafkia *et al*., Operation of a TCA cycle subnetwork in the mammalian nucleus. *Sci. Adv.* **8**, eabq5206 (2022). doi: 10.1126/sciadv.abq5206; pmid: 36044572
48. X. Liu *et al*., The existence of a nonclassical TCA cycle in the nucleus that wires the metabolic-epigenetic circuitry. *Signal Transduct. Target. Ther.* **6**, 375 (2021). doi: 10.1038/s41392-021-00774-2; pmid: 34728602
49. M. Xiao *et al*., Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012). doi: 10.1101/gad.191056.112; pmid: 22677546
50. D. Kalderon, B. L. Roberts, W. D. Richardson, A. E. Smith, A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509 (1984). doi: 10.1016/0092-8674(84)90457-4; pmid: 6096007
51. C. Dubos *et al*., MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581 (2010). doi: 10.1016/j.tplants.2010.06.005; pmid: 20674465
52. C. Li *et al*., Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in *Arabidopsis*. *Nat. Genet.* **48**, 687–693 (2016). doi: 10.1038/ng.3555; pmid: 27111034
53. G. C. Russell, J. R. Guest, Sequence similarities within the family of dihydrolipoamide acyltransferases and discovery of a previously unidentified fungal enzyme. *Biochim. Biophys. Acta* **1076**, 225–232 (1991). doi: 10.1016/0167-4838(91)90271-Z; pmid: 1825611
54. X. Cui *et al*., REF6 recognizes a specific DNA sequence to demethylate H3K27me3 and regulate organ boundary formation in *Arabidopsis*. *Nat. Genet.* **48**, 694–699 (2016). doi: 10.1038/ng.3556; pmid: 27111035
55. W. Yang, D. Jiang, J. Jiang, Y. He, A plant-specific histone H3 lysine 4 demethylase represses the floral transition in *Arabidopsis*. *Plant J.* **62**, 663–673 (2010). doi: 10.1111/j.1365-313X.2010.04182.x; pmid: 20202164
56. M. Lei *et al*., Ethylene signalling is involved in regulation of phosphate starvation-induced gene expression and production of acid phosphatases and anthocyanin in *Arabidopsis*. *New Phytol.* **189**, 1084–1095 (2011). doi: 10.1111/j.1469-8137.2010.03555.x; pmid: 21118263
57. F. Ma, L. J. Jazmin, J. D. Young, D. K. Allen, Isotopically nonstationary 13C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16967–16972 (2014). doi: 10.1073/pnas.1319485111; pmid: 25368168
58. Q. Tang *et al*., An integrated isotopic labeling and freeze sampling apparatus (ILSA) to support sampling leaf metabolomics at a centi-second scale. *Plant Methods* **18**, 97 (2022). doi: 10.1186/s13007-022-00926-7; pmid: 35907895
59. M. D. Curtis, U. Grossniklaus, A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiol.* **133**, 462–469 (2003). doi: 10.1104/pp.103.027979; pmid: 14555774
60. M. Karimi, B. De Meyer, P. Hilson, Modular cloning in plant cells. *Trends Plant Sci.* **10**, 103–105 (2005). doi: 10.1016/j.tplants.2005.01.008; pmid: 15749466
61. K. Wang *et al*., *EAR1* negatively regulates ABA signaling by enhancing 2C protein phosphatase activity. *Plant Cell* **30**, 815–834 (2018). doi: 10.1105/tpc.17.00875; pmid: 29618630
62. I. A. Sparkes, J. Runions, A. Kearns, C. Hawes, Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* **1**, 2019–2025 (2006). doi: 10.1038/nprot.2006.286; pmid: 17487191
63. Y. Wang, X. Gu, W. Yuan, R. J. Schmitz, Y. He, Photoperiodic control of the floral transition through a distinct polycomb repressive complex. *Dev. Cell* **28**, 727–736 (2014). doi: 10.1016/j.devcel.2014.01.029; pmid: 24613395
64. I. Nigi, L. Fairall, J. W. R. Schwabe, Expression and purification of protein complexes suitable for structural studies using mammalian HEK 293F cells. *Curr. Protoc. Protein Sci.* **90**, 5.28.1–5.28.16 (2017). doi: 10.1002/cpps.44; pmid: 29091272
65. P. A. Longo, J. M. Kavran, M. S. Kim, D. J. Leahy, Transient mammalian cell transfection with polyethylenimine (PEI). *Methods Enzymol.* **529**, 227–240 (2013). doi: 10.1016/B978-0-12-418687-3.00018-5; pmid: 24011049
66. H. He, J. Wang, T. Liu, UV-Induced RPA1 acetylation promotes nucleotide excision repair. *Cell Rep.* **20**, 2010–2025 (2017). doi: 10.1016/j.celrep.2017.08.016; pmid: 28854355
67. F. Huang, W. Yuan, S. Tian, Q. Zheng, Y. He, *SIN3 LIKE* genes mediate long-day induction of flowering but inhibit the floral transition in short days through histone deacetylation in *Arabidopsis*. *Plant J.* **100**, 101–113 (2019). doi: 10.1111/tpj.14430; pmid: 31168864
68. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi: 10.1093/bioinformatics/btp616; pmid: 19910308
69. M. Louwers, E. Splinter, R. van Driel, W. de Laat, M. Stam, Studying physical chromatin interactions in plants using Chromosome Conformation Capture (3C). *Nat. Protoc.* **4**, 1216–1229 (2009). doi: 10.1038/nprot.2009.113; pmid: 19644461
70. Z. Li *et al*., JASMONATE-ZIM DOMAIN proteins engage Polycomb chromatin modifiers to modulate Jasmonate signaling in *Arabidopsis*. *Mol. Plant* **14**, 732–747 (2021). doi: 10.1016/j.molp.2021.03.001; pmid: 33676023
71. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). doi: 10.1038/nmeth.1923; pmid: 22388286
72. H. Li *et al*., The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009). doi: 10.1093/bioinformatics/btp352; pmid: 19505943
73. Y. Zhang *et al*., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
74. G. Yu, L. G. Wang, Q. Y. He, ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015). doi: 10.1093/bioinformatics/btv145; pmid: 25765347
75. D. Li, S. Hsu, D. Purushotham, R. L. Sears, T. Wang, WashU Epigenome Browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019). doi: 10.1093/nar/gkz348; pmid: 31165883
76. F. Ramírez *et al*., deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). doi: 10.1093/nar/gkw257; pmid: 27079975

ACKNOWLEDGMENTS

We thank L. Peng and H. Zhang (Shanghai Center for Plant Stress Biology, CAS) for help with the bioinformatic analysis and for providing *IBM1:Flag* seeds, respectively; Q. Liu and J. Yang (Peking University, China) for critically reading the manuscript; J. Li (Hainan University, China) for discussion; members of the He laboratory for assistance; and the Protein Preparation and Characterization Core Facility of Tsinghua University Branch of China National Center for Protein Sciences Beijing for assistance with mammalian protein expression. **Funding:** This work was supported in part by the National Natural Science Foundation of China (grant nos. 31830049 and 31721001 to Y.H.), the Peking-Tsinghua Center for Life Sciences, and CAS. **Author contributions:** Y.H. conceived the study. F.H.,

X.L., and Y.O. designed and performed most of the experiments. Z.G. conducted bioinformatic analysis. Q.T. and Z.C. performed part of the experiments. F.H., X.L., Y.O., Y.H., Q.T., and Z.C. analyzed experimental data. Y.H. and X.Z. supervised the research. Y.H. wrote the paper with the help from F.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The raw data from ChIP-seq have been deposited in the Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information (NCBI), under the following accession numbers: GSE182411 (H3K27me3 ChIP-seq) and GSE216563 (OGDH1 ChIP-seq). The raw data from RNA-seq have been deposited at the NCBI Sequence Read Archive under the accession number of PRJNA894352. The raw data from H3K4me3 ChIP-seq are available at NCBI under the accession number GSE185121. All other data are available in the main text or supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf8822
Figs. S1 to S12
Tables S1 to S5
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 18 November 2022; accepted 19 May 2023
10.1126/science.adf8822

RESEARCH ARTICLE SUMMARY

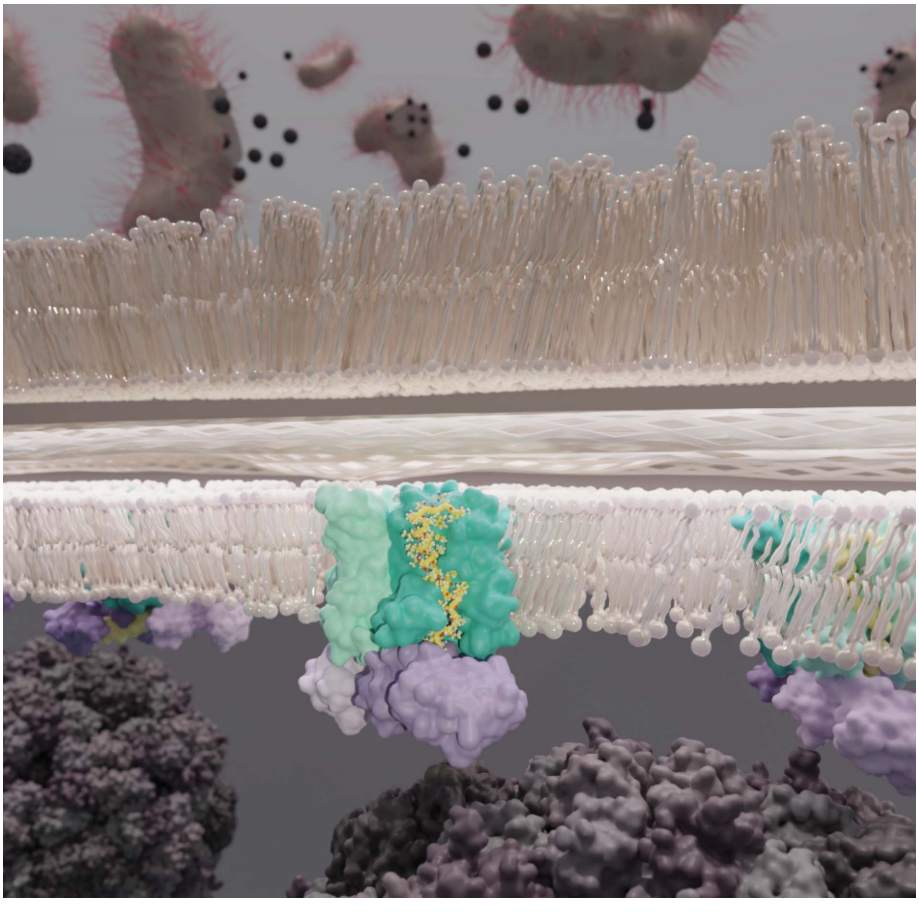
ANTIMICROBIAL PROTEINS

The mechanism of the phage-encoded protein antibiotic from ΦX174

Anna K. Orta†, Nadia Riera†, Yancheng E. Li, Shiho Tanaka, Hyun Gi Yun, Lada Klaic, William M. Clemons Jr.*

INTRODUCTION: As part of their life cycle, viruses must escape from their hosts. For bacteriophages, this process requires breaching the bacterial cell wall and the peptidoglycan layer. Small, single-strand DNA and RNA phages have evolved single-gene lysis proteins that disrupt peptidoglycan biosynthesis and trigger cell lysis, a mechanism also used by some antibiotics. The best-studied of these is protein E from the historically important phage ΦX174, a 91–amino acid protein that contains a conserved N-terminal transmembrane helix and an extended cytoplasmic C-terminus.

RATIONALE: Mutagenesis studies have demonstrated that the protein E transmembrane domain is sufficient for lysis, with functional mutants mapping to one face of the single transmembrane helix. The strongest phenotypes were observed for conserved prolines in this domain of protein E, with one proline being essential for lysis. Wild-type protein E requires the constitutively expressed bacterial chaperone SlyD (sensitivity to lysis D), which contains a prolyl isomerase domain and a chaperone domain. Previous work had identified mutants in MraY that confer resistance to protein E–mediated lysis. MraY catalyzes the synthesis of a peptidoglycan precursor from a soluble nucleoside-sugar-peptide and a phospholipid substrate. Although many functional questions remain about MraY, here we sought to understand how a phage protein could specifically inhibit this key enzyme and what role SlyD performed in the process.



The escape of ΦX174 from its bacterial host. In the membrane is the YES complex [*E. coli* enzyme MraY (cyan), phage protein E (yellow), and *E. coli* chaperone SlyD (purple)] where protein E disrupts peptidoglycan synthesis by inhibiting MraY and allowing breaching of the cell wall (tan). Assembled ΦX174 phage particles are shown in gray. In the background are bacterial cells that are lysing at their septal division point.

RESULTS: We established that phage protein E forms a stable complex with the *Escherichia coli* proteins MraY and SlyD, designated the YES complex (MraY, protein E, SlyD). After solubilizing in detergent, we determined structures of the YES complex by single-particle electron cryo–microscopy. In our structures, the MraY dimer adopts a back-to-back orientation with the membrane-exposed active sites facing away from each other. Two SlyD molecules can be fit into the density on the cytoplasmic side. The two MraY and SlyD chains are bridged by two protein E molecules. The transmembrane helix of protein E occupies a groove on MraY corresponding to the putative binding site of the lipid substrate. At the cytoplasmic surface, the transmembrane domain of protein E bends to cross the active site, followed by an α-helix that extends to the chaperone domain of SlyD. The C-terminus of protein E continues through a peptide-binding groove in the second SlyD and into the prolyl isomerase active site. The consequence is that protein E inhibits MraY by blocking lipid access to the active site. Previous protein E mutant phenotypes are explained by the structure including the essential proline that allows for a kink in the transmembrane helix. For MraY, we can resolve the full chain including the conserved loop between TM1 and TM2. We show that the N terminus of MraY forms an α-helical stack with TM2 and identify ordered lipids on the protein surface. Finally, we provide evidence that the role of SlyD is to stabilize the complex.

CONCLUSION: Protein E directly inhibits MraY by obstructing the MraY active site in a stable complex with SlyD. This structure resolves key questions about how the model phage ΦX174 kills bacteria and escapes the cell. In the ΦX174 genome, the gene encoding for protein E is evolutionarily constrained by gene D, in which it is embedded. The YES complex provides a route for the rational design of protein E beyond this gene D restraint. Single-gene lysis proteins, like protein E, serve as useful models for the development of antibacterial therapies. ■

The list of author affiliations is available in the full article.
*Corresponding author. Email: clemons@caltech.edu
†These authors contributed equally to this work.
Cite this article as A. K. Orta *et al.*, *Science* **381**, eadg9091 (2023). DOI: 10.1126/science.adg9091

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adg9091

RESEARCH ARTICLE

ANTIMICROBIAL PROTEINS

The mechanism of the phage-encoded protein antibiotic from ΦX174

Anna K. Orta[1]†, Nadia Riera[1]†, Yancheng E. Li[1], Shiho Tanaka[1], Hyun Gi Yun[1], Lada Klaic[1], William M. Clemons Jr.[1]*

The historically important phage ΦX174 kills its host bacteria by encoding a 91-residue protein antibiotic called protein E. Using single-particle electron cryo–microscopy, we demonstrate that protein E bridges two bacterial proteins to form the transmembrane YES complex [MraY, protein E, sensitivity to lysis D (SlyD)]. Protein E inhibits peptidoglycan biosynthesis by obstructing the MraY active site leading to loss of lipid I production. We experimentally validate this result for two different viral species, providing a clear model for bacterial lysis and unifying previous experimental data. Additionally, we characterize the *Escherichia coli* MraY structure—revealing features of this essential enzyme—and the structure of the chaperone SlyD bound to a protein. Our structures provide insights into the mechanism of phage-mediated lysis and for structure-based design of phage therapeutics.

The full realization of phage therapy as a solution to the antimicrobial resistance problem requires a fundamental understanding of the mechanisms viruses use to kill their host (*1*–*4*). Nearly 100 years ago, the first medical application of phages to treat infections used a cocktail containing the historically important phage ΦX174 (*5*). A rich source of critical discoveries in molecular biology (*6*–*9*), ΦX174 is a member of the *Bullavirinae* subfamily within *Microviridae*. It is found broadly in environments that contain coliform bacteria, such as the human gut (*10*), with *Escherichia coli* as its primary host. The 1977 publication of the ΦX174 single-stranded 5.4-kilobase DNA genome was a milestone for genomics, revealing 11 open reading frames (ORFs) (*11*). Most notable was the gene for lysis, *E*, that is embedded within the ORF of the scaffolding gene *D* (*12*). Expression of *E* alone is sufficient to kill bacteria (*13*). Although a variety of lysis mechanisms have been proposed (*14*–*16*), the most likely is that the product of *E*, protein E, disrupts peptidoglycan (PG) biosynthesis, culminating in a breach in the cell wall. In particular, protein E was demonstrated to directly inhibit the integral membrane enzyme MraY (*16*) dependent on the cytoplasmic chaperone SlyD (sensitivity to lysis D) (*17*, *18*).

Phage-derived single-gene lysis (SGL) proteins, such as protein E, trigger cell lysis by inhibiting PG biosynthesis [reviewed in (*19*)] providing a route for killing bacteria. Protein E includes a conserved N-terminal transmembrane domain (TMD), a cytoplasmic C-terminal domain with conserved positively charged residues in a predicted amphipathic helix, and an unstructured tail (Fig. 1A and fig. S1). Fusions of globular proteins to the C terminus of the TMD have demonstrated that the TMD is sufficient for lysis (*17*, *20*, *21*). Extensive mutational analysis identified key residues in protein E including the essential proline 21 (P21) (*17*, *20*–*23*). The mechanism through which SlyD and MraY work with protein E to facilitate rupture of the cell wall remains unknown. MraY, an important target for antimicrobial drug discovery [reviewed in (*24*)], catalyzes the transfer of a phospho-MurNAc-pentapeptide from "Park's nucleotide" (UDP-MurNAc-pentapeptide) onto the phosphate group of the 55 carbon polyisoprenyl-phosphate ($C_{55}P$) generating lipid I. X-ray crystal structures of MraY in detergent from the thermophilic Gram-negative *Aquifex aeolicus* (*Aa*MraY) (*25*, *26*) or the Gram-positive *Enterocloster bolteae* (*Eb*MraY) (*27*) revealed a conserved homodimeric structure, with ten TMDs per subunit. The structures localized the active site in a vestibule on the cytoplasmic side of a membrane-exposed cleft formed by TMDs 4, 5, and 9, which is the predicted access site for the lipid substrate (*28*). The constitutively expressed but nonessential (*29*) metallochaperone (*30*) SlyD contains two conserved core domains: an FK506-binding protein (FKBP) prolyl-isomerase domain and an insertion-in-flap (IF) domain that binds unfolded peptides by β-augmentation (*18*, *31*–*33*). The disordered SlyD C terminus (*31*), although important for metal binding, can be deleted without affecting chaperone activity (*34*).

In the present study, we resolve the mechanism of peptidoglycan biosynthesis inhibition by protein E. We determined the structure of the dimeric heterotrimer YES complex (*Ec*MraY, viral protein E, *Ec*SlyD) revealing that protein E physically blocks the lipid substrate from accessing the MraY active site. This work provides mechanistic insight into all three proteins and suggests a path toward development of antibacterial agents.

The structure of the YES complex

We used the protein E sequence from either the original phage ΦX174 or the shorter protein E isoform from phage ID21 (91 and 76 residues respectively), both from within *Bullavirinae* (Fig. 1A and fig. S1). We coexpressed an affinity-tagged protein $E_{ΦX174}$ and wild-type (WT) *Ec*MraY. After purification, a stable complex formed a single peak during size-exclusion chromatography (SEC) but resolved into four bands on a gel, two of which corresponded to the endogenous *Ec*SlyD (fig. S2A). A SlyD variant with the disordered C terminus removed rescued lysis activity in an *E. coli* Δ*slyD* strain (fig. S2, B to D) and ran as a single band on a gel (Fig. 1B) (*18*, *30*). This truncation, $SlyD_{154}$, was used for subsequent work except where noted.

The two protein E isoforms induced lysis at similar efficiencies when expressed in a WT *E. coli* strain (Fig. 1C). For purification, all three genes in the YES complex (WT *Ec*MraY, protein E, and truncated *Ec*SlyD) were recombinantly expressed together in the Δ*slyD* strain with a C-terminal affinity tag on protein E. The complex was extracted in detergent and ran as a single peak by SEC with all three proteins in an apparent stoichiometric complex (Fig. 1B). Structures for both the $YES_{ΦX174}$ and YES_{ID21} complexes were solved using single-particle electron cryo–microscopy (cryo-EM) (Fig. 1D and figs. S3 and S4). The final density maps were obtained following several rounds of data processing with heterogeneous and homogeneous refinements (figs. S3 and S4). The final overall masked resolution was 3.5 Å for the YES_{ID21} complex and 3.6 Å for the $YES_{ΦX174}$ complex (fig. S5). Statistics are provided in table S2. For both structures, the resolution was higher for regions in and adjacent to the membrane (figs. S3 and S4). We unambiguously built 90% of the protein residues in the complex. Sequence differences between the protein E variants are visible in the density (fig. S5), generally exposed on the surface of the complex. The YES_{ID21} complex is used as the reference structure, except where noted.

Within the density map, we could clearly distinguish two copies of each member of the YES complex (six separate proteins). When contoured to remove the detergent micelle, densities for 22 TMDs are clear, 20 of which are accounted for by the *Ec*MraY dimer (Fig. 1, D and E). Most of the cytoplasmic density can be accounted for by two SlyD molecules, which is the most flexible region of the complex (fig. S6D). The remaining protein density corresponds to two protein E molecules

[1]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA.
*Corresponding author. Email: clemons@caltech.edu
†These authors contributed equally to this work.

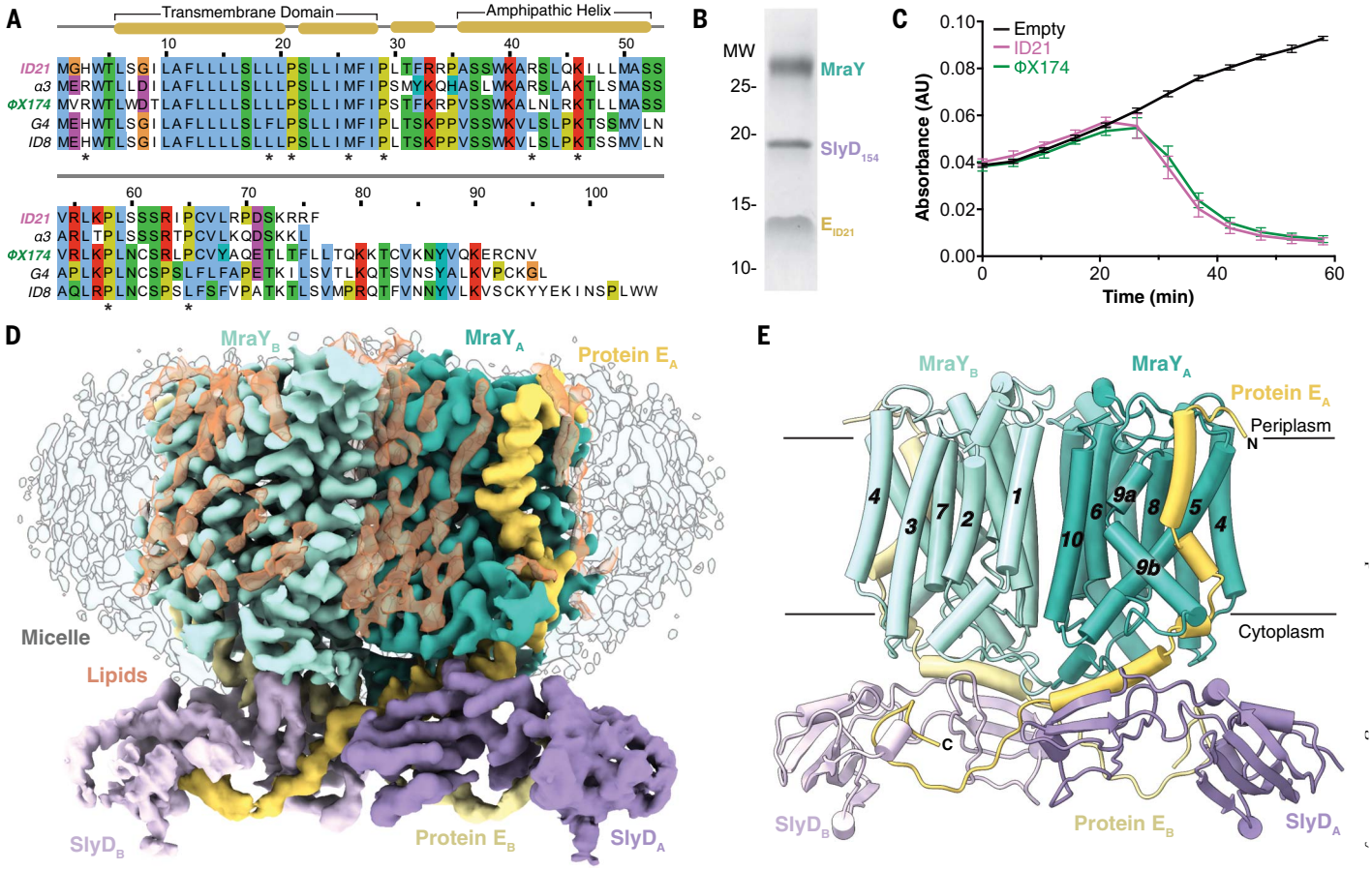


Fig. 1. The structure of the YES complex. (**A**) An alignment of a representative subset of protein E isoforms. Residue coloring is based on ClustalW. Secondary structure elements are shown above the sequence. Sequences are ordered as in fig. S1. Residues highlighted with a (*) at the bottom of the alignment are discussed in the text. (**B**) SDS-PAGE gel of the purified YES_{ID21} complex. (**C**) A lysis assay for protein E expression. Cells containing either empty vector (black line) or the protein E genes for ID21 (pink line) or ΦX174 (green line) were induced at time 0 and the absorbance at 600 nm was monitored over time. Error bars represent the standard deviation derived from $n = 3$. (**D**) Overview density maps of the YES_{ID21} complex viewed in the plane of the membrane. The detergent micelle is highlighted by gray coloration. The higher contoured map shows the six components of the twofold complex with density for *E. coli* MraY (cyan), protein E (yellow), and *E. coli* SlyD (purple) highlighted. The pairs of each protein are distinguished as the B-subunit is light purple. The general coloring scheme is maintained throughout the figures. Density that is likely lipid is shown in transparent orange. (**E**) Illustrated representation of the YES_{ID21} complex oriented and colored as in (D). Foreground TMDs for the MraYs are numbered and the bilayer is represented by lines. N terminus and C terminus of protein E_A are labeled.

that each contain a kinked TMD and a soluble domain that bridges between MraY and SlyD. Densities for lipids are visible around the membrane-exposed surface of the MraY dimer (Fig. 1D).

For protein E, starting at the N terminus in the periplasm, the TMD binds in the groove formed between TM5 and TM9 of MraY ending on the cytoplasmic side where it makes a sharp turn into the active site pocket (Figs. 1E and 2A). The TMD is followed by an amphipathic helix that crosses the active site, parallel to the membrane, presenting a positive face toward MraY and a hydrophobic face toward a SlyD IF-domain. The C-terminal residues adopt an extended conformation that primarily interacts with the second SlyD. This results in a crossover point between the two protein Es with each contacting both SlyDs. Overall, the dimeric complex (two of each of the heterotrimers) has a near twofold symmetry perpendicular to the plane of the membrane. The membrane and periplasmic facing regions overlay perfectly but the symmetry is broken at the cytoplasmic face where the C-termini of the two protein E molecules cross each other at different residues and the SlyDs adopt slightly different orientations (Fig. 1E and fig. S6). Following toward the end of protein E, we can see continuous backbone density that positions proline 65 in the active site of the FKBP domain. Beyond that, the density is insufficient to resolve the sequence and we see little difference between ΦX174 and ID21 (fig. S5).

The interaction of protein E with MraY

Functional studies have consistently revealed the requirement for a proline at position 21 (*21*, *22*, *35*). Our structure allows an elegant explanation for this requirement. The protein E TMD binds in the cleft of MraY that is defined by the angled TM9b (Fig. 2A). The proline at position 21 breaks the hydrogen bonding and introduces a kink that allows bending of the TMD around TM9b following the groove to the active site. Mutation of P21 would result in loss of the kink favoring a straight TMD that could not bind in the groove. Residue P29, also completely conserved (Fig. 1A and fig. S1), creates a second bend that completes the wrap around TM9b and a mutation at this position results in delayed lysis onset (*21*, *22*). Additional alanine mutations in the TMD identified residues that result in delayed lysis onset, postulated to decrease the binding affinity of protein E to MraY (*21*). In the present structure, most of these residues (L19, L20, L23, and M26) (Fig. 2C) make direct contact with MraY. The exception, F27, appears

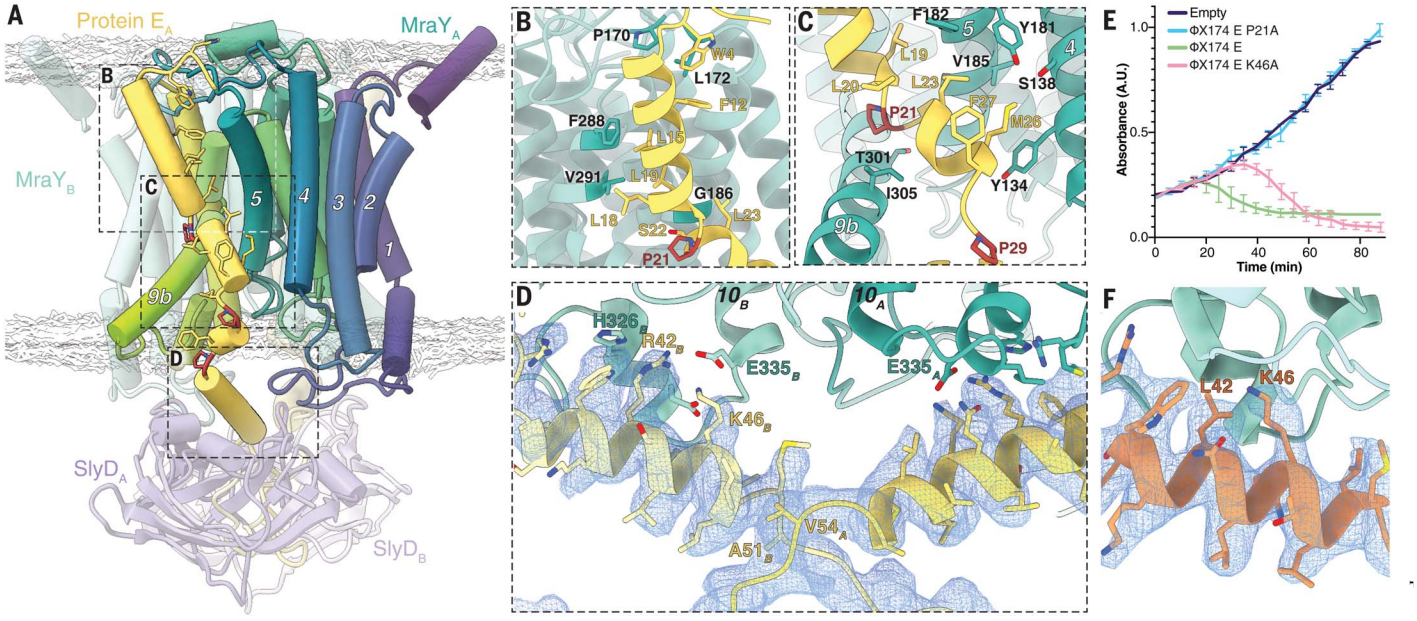


Fig. 2. The interaction of protein E with *Ec*MraY. (**A**) Illustration of the YES$_{ID21}$ complex as in Fig. 1E with a 90° rotation except that MraY$_A$ is color-ramped using the Viridis palette that goes from purple to yellow for the N- to C-terminus, respectively. Prolines are shown in red. The bilayer is modeled in white. The B-subunits and SlyDs are faded. Boxes indicate regions highlighted in panels (**B** to **D**). Side chains for protein E residues that contact MraY are shown as sticks. (B) as in (A) with positions of protein E resistance mutants in MraY highlighted (dark cyan). (C) as in (A) with residues at the interface highlighted as sticks. (D) The region where the two protein E molecules cross highlighting the asymmetry colored as in (A) with interacting residues shown as sticks. Density for the amphipathic helix of protein E shown as a blue mesh. SlyDs are removed for clarity. (**E**) Lysis assay of expressed protein E$_{\Phi X174}$ variants. Error bars represent the standard deviation derived from n = 3. (**F**) is similar to (D) for protein E$_{\Phi X174}$.

to sterically position M26 into a tight interaction with Y134 in MraY, which is conserved in most Gram-negative bacteria (fig. S7).

At the cytoplasmic interface, protein E residues A36 through M50 form the amphipathic helix spanning the width of an MraY subunit. The hydrophilic face of this helix orients toward the membrane in the MraY active site. The helix contains conserved positively charged residues that interact with conserved negatively charged residues in MraY (Fig. 2D). An example is the K46 salt bridge where, in our lysis assay, a K46A mutation resulted in delayed lysis onset (Fig. 2E). Here, isoform differences can be visualized. For example, position 42 is a leucine in ΦX174 whereas in ID21 it is an arginine that forms a salt bridge with E335 in MraY (Fig. 1). The helix ends at the crossover between the two protein Es at residue V54 in protein E$_A$ and A51 in protein E$_B$ resulting in differing interactions with residues in MraY, such as the essential H326 (Fig. 2D).

The Epos (plaques on Δ*slyD*) mutations, R3H and L19F, allow phage propagation in a Δ*slyD* background (*23*). The R3H mutant of protein E in ΦX174 results in a silent mutation in protein D and is found native to other species such as ID21 (fig. S8). Previous work reported that this variant resulted in higher levels of protein E in the membrane (*23*). In the present structure (fig. S5, G and H), this residue does not make specific contact with MraY. It is likely that the loss of a positive charge in the periplasm favors a higher percentage of protein E molecules that are correctly inserted in the membrane as a result of the positive-inside rule. L19F does not result in higher levels of protein E (*21*) and likely hastens lysis onset by having a higher affinity to MraY. Another phenylalanine mutation at the interface, L23F, also hastens lysis onset likely by higher affinity, although this is not a general rule as other leucine to phenylalanine mutants did not affect lysis (*21*). Both L19 and L23 are near the conserved F182 in MraY and may add additional stability through aromatic π interactions (Fig. 2C). The opposite mutation, phenylalanine to leucine, can show loss of binding. For example, the F288L mutation in MraY (*16*) is located at the interface with protein E and results in a loss of lysis, most likely due to lower affinity.

All the mutations in MraY that allow resistance to protein E–mediated lysis (P170L, ΔL172, G186S, F288L, V291M) (*16*, *35*) are at the interface with protein E (Fig. 2B). The mutant F288L, as noted above, lowers affinity to protein E. Residue G186 is at the nearest approach between the two proteins and a mutation to serine would prevent protein E binding. The mutant V291M lies directly at the interface near L19 in protein E, although a specific effect for this mutation is not clear. Finally, P170L and ΔL172 are located within periplasmic loop 4-5, which interacts with the conserved N terminus of protein E. Although more resistant to lysis, these two mutants are predicted to still bind protein E, albeit with lower affinity (*35*).

The mechanism of inhibition of MraY by protein E

The YES complex structure allows us to propose a simple mechanism for inhibition of MraY by protein E. Superposition with previous MraY inhibitor complexes (*26*, *27*) on the YES complex (Fig. 3) shows that the predicted path of the polyisoprenyl chain of C$_{55}$P is the groove formed between TM5 and TM9, which is occluded by protein E. Therefore, one mechanism of inhibition is that protein E prevents access of the lipid substrate to the active site (Fig. 3A). Protein E is a noncompetitive inhibitor of Park's nucleotide (*36*) and, accordingly, the pocket that binds the nucleoside is fully accessible in the present structure (Fig. 3B). Loop 9-10 in MraY contains catalytic histidines that must move toward the binding pocket to facilitate catalysis by completing the active site (*28*). The cytoplasmic helix of protein E separates loop 9-10 from the rest of the active site, blocking this transition and providing a second mechanism of inhibition. Overall, protein E blocks access of the lipid substrate and prevents formation of the active site upon substrate binding.

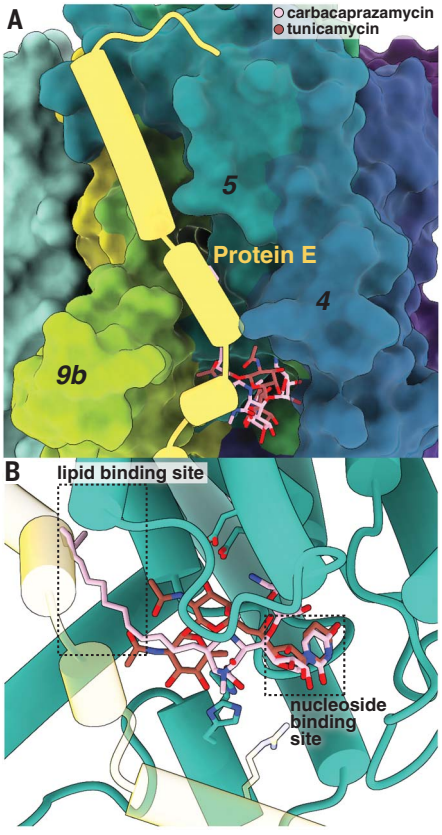


Fig. 3. The mechanism of inhibition by protein E. (**A**) Accessible surface of MraY (colored as in Fig. 2A) viewed from the cytoplasm looking toward the active site cleft. Protein E is shown as an illustration. The structures for the inhibitors tunicamycin (maroon) and carbacaprazamycin (pink) are shown as sticks based on a structural alignment to *Ec*MraY$_A$ with their respective complex structures [PDBID: 6OYH (*26*) and 5JNQ (*27*)]. (**B**) Similar to (A) from a slightly different angle highlighting the catalytic pocket. MraY (dark cyan) is shown as an illustration. The two substrate binding sites are highlighted by dashed boxes. Predicted catalytic residues in MraY are shown as sticks.

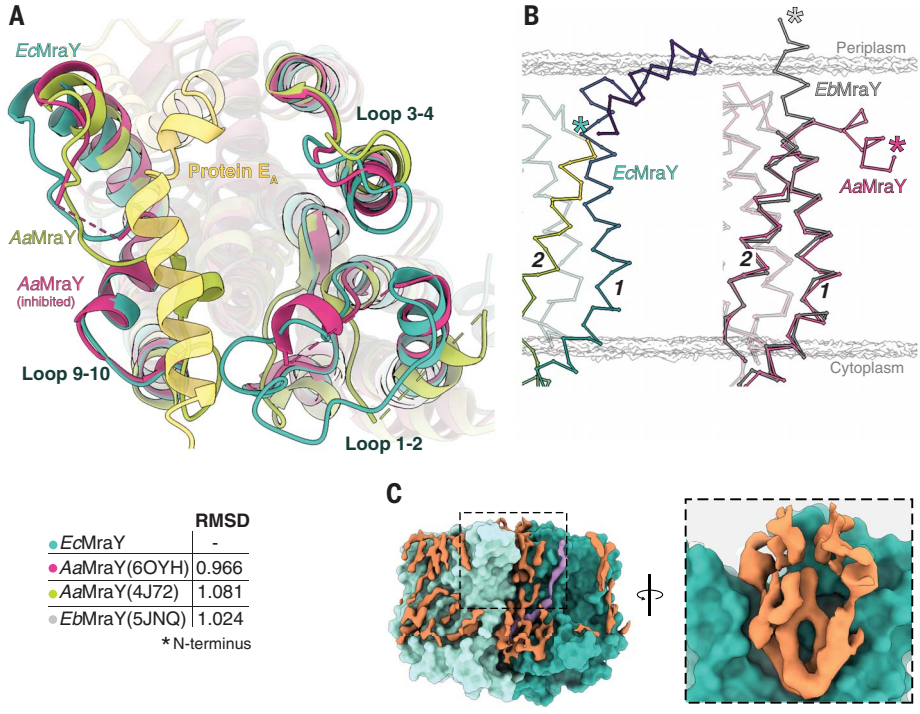


Fig. 4. Features observed in the EM structure of *E. coli* MraY. (**A**) Cytoplasmic view of a structural alignment of *Ec*MraY (dark cyan) against uninhibited *Aa*MraY [green, PDBID:4J72 (*25*)] and carbacaprazamycin inhibited *Aa*MraY [pink, PDBID:6OYH (*26*)]. RMSDs to monomer A of *Ec*MraY are shown. The color scheme for the various MraY crystal structures is used throughout the figures. (**B**) A view in the plane of the membrane showing the region that includes TM1 and TM2 in backbone ribbons. (Left) The *Ec*MraY structure colored from the N terminus through TM2 in the Viridis color scheme. (Right) The inhibited *Aa*MraY structure from (A) and the *Eb*MraY structure [gray, PDBID:5JNQ (*27*)]. Each structure is aligned to *Ec*MraY. The location of each N terminus is indicated by an asterisk. (**C**) Accessible surface of *Ec*MraY (cyan) and unmodeled densities (orange) that are likely lipids or detergent as well as a putative C55P molecule (purple, see below). The inset is a view of the periplasmic cavity viewed from a removed monomer.

Key structural features of MraY

Although there are prior crystallographic MraY structures (*25*–*27*), our YES complex structure reveals *E. coli* MraY in a distinct structural context (fig. S9). There is agreement with the prior structures when comparing MraY monomers with backbone RMSDs around 1 Å (Fig. 4A and fig. S10). We can model previously disordered regions of MraY including all the cytoplasmic loops that enclose the active site. Loop 1-2 likely adopts distinct conformations during the catalytic cycle and indeed has two slightly different conformations in our dimer (fig. S6C). Loop 9-10 adopts a conformation that is similar to the reported small-molecule inhibitor co-crystal structures of MraY (*26*, *27*, *37*) but distinct from the apoprotein structure (Fig. 4 and fig. S10). This conformation of loop 9-10 may be a general feature of MraY complexed with inhibitors.

In prior crystal structures, the N terminus of MraY begins at either TM1 or is a helix that projects away from the structure in an orientation incompatible with the bilayer (Fig. 4B). In the YES complex, the N-terminal end of the first helix (NTH) hydrogen-bonds to the C-terminal end of TM2, effectively forming a helical stacking structure (Fig. 4B and fig. S9C). A multiple sequence alignment across bacteria shows that the NTH in MraY is conserved across gram-negative bacteria but missing in Gram positives (figs. S7 and S11). Although this feature is not found in the crystal structures, AlphaFold (*38*) predicts the NTH stacking for *E. coli* and other Gram-negative bacteria with slight differences in hydrogen bonding and orientation relative to the cryo-EM structure (fig. S11). For the *Aa*MraY structures (*25*, *26*), the positioning of the N terminus is likely a product of crystallization, as AlphaFold predicts the NTH stacks for this and the related *Hydrogenivirga* species (fig. S11). For Gram-positive bacteria, both the *Eb*MraY and predicted structures lack the NTH (*27*) (Fig. 4B and fig. S11).

Protein E as a general antibacterial protein

Expression of protein E leads to lysis that results in a subset of killed bacteria becoming "ghosts"—essentially empty cell walls (*15*). Ghosts have been used in a variety of contexts including vaccine development [reviewed in (*39*)]. For ghosts to be made, protein E must inhibit the native MraY of the target bacteria. Although phage ΦX174 is restricted to *E. coli*, many bacteria have been probed for ghost formation by expression of protein E. All Gram-negative bacteria tested (fig. S11A) resulted in the formation of ghosts [reviewed in (*40*)]. By contrast, protein E was unable to cause lysis in the Gram-positive *Staphylococcus carnosus* (*41*). Expression of the *Bacillus subtilis* MraY in *E. coli* prevented cell lysis, suggesting that it has a low affinity for protein E (*35*). Comparing the residues in *Ec*MraY that contact protein E to those in Gram-positive species, there are no single residue differences that easily explain the inability to inhibit MraY from gram-positive species. There are a few residues that may play a role—such as P170 in *Ec*MraY which confers resistance when mutated to leucine—that are absent in Gram-positive MraYs (fig. S7).

*Ec*MraY residue Y134, which forms a bridge with protein E M26 in our model (Fig. 2C), is also missing in Gram-positive species (fig. S7).

Lipids bound to the YES complex

The YES complex was solubilized from its native environment and, for the YES_{ID21} complex, the final detergent solution was supplemented with *E. coli* lipids. We observe lipid densities around the membrane surface of MraY (Fig. 1D and fig. S12, A and B). Prior work supports a functional role for anionic phospholipids with MraY including stabilizing the dimer (*42*–*44*). We observe substantial lipid density near the dimer interface with features consistent with glycerophospholipids, although not all density can be clearly assigned (fig. S12). As in previous structures (*25*, *27*), there is a hydrophobic periplasmic cavity within the MraY dimer that contains unexplained density (Fig. 4C). Although this density has clear structure, we are unable to fit typical *E. coli* phospholipids.

Recently, a study using native mass spectrometry and molecular dynamics identified a binding site for $C_{55}P$ at the dimer interface with the phosphate headgroup predicted to form a salt bridge to R341 in *Ec*MraY (*44*). In our structure, we observed a long tube of density at the dimer interface (fig. S12A, purple) that ends at R341 and is consistent with $C_{55}P$ (fig. S12C). Although we cannot confidently identify this lipid, the density would be consistent with $C_{55}P$, although the role for this binding site remains to be determined.

The role of SlyD in protein E–mediated lysis

The amphipathic helix of protein E bridges the two *E. coli* proteins, MraY and SlyD, which make no specific contact with each other (Fig. 2A). The IF domain sits on the hydrophobic face of the protein E helix (Fig. 5A) and contacts the extended C terminus of the opposing protein E, which then extends to bind the FKBP domain (Fig. 5B). This results in a bowtie-like interaction with each SlyD binding both protein E–soluble domains (Fig. 1E). There are no contacts between the two SlyD molecules (fig. S6D). These protein E interfaces validate structural studies of SlyD where peptides bind to each of the interfaces seen here including β-augmentation in a groove in the IF domain (Fig. 5, C and D) (*32*). The IF domain binding to an α-helix may be an important chaperoning interaction. The FKBP domain is well-ordered but adopts a range of orientations in our particles (fig. S13) and consequently has the lowest resolution (fig. S3). This flexibility relative to the IF domain is consistent with the NMR structure of SlyD where the orientations of the two domains are not constrained by each other (*31*). The only contact of the FKBP domain to the rest of the YES complex is the flexible linker to the IF domain and binding to the extended C terminus of protein E. We observe continuous backbone density that allows us to place P65 at the FKBP active site. As seen before for some SlyD-bound peptides (*32*), protein E binding to the FKBP domain adopts a noncanonical orientation. Protein E P65 is not completely conserved (fig. S1), although other species have a proline at residue 63 which could reach the active site with additional tilting of the FKBP domain. A proline near this position may help to localize the chaperone to the complex and facilitate assembly.

The lack of contacts between SlyD and MraY suggest that the soluble domain of protein E alone could form a complex with SlyD. We coexpressed *Ec*SlyD with N-terminal truncations of protein E from either ID21 (residues 33 to 76) or the shortest isoform α3 (residues 33 to 75). Both form complexes that could be purified by an affinity tag on the soluble domain (fig. S2E). These observations point to a high-affinity interaction between SlyD and the C-terminal domain of protein E.

Protein E is unstable in the absence of SlyD (*18*) and is rapidly degraded (*23*). It has been speculated that prolyl-isomerization is central to the lysis mechanism (*22*), but this idea is not supported by available evidence. Nonproline mutations in protein E can rescue lysis in a Δ*slyD* background (*18*, *45*), as well as completely replace the cytoplasmic domain of protein E with an unrelated globular protein (*20*, *21*, *45*). We also performed our lysis assay with several SlyD variants in the Δ*slyD* strain (fig. S2C). As before, protein $E_{\Phi X174}$ was unable to promote lysis in the absence of SlyD, however lysis could be rescued by expression of either *Ec*SlyD or $EcSlyD_{154}$. *Thermus thermophilus* (*Tt*) SlyD has high structural homology to *Ec*SlyD (Fig. 5D and fig. S2, B, F, and G) and also rescued lysis in the Δ*slyD* strain (fig. S2C). We generated an *Ec*SlyD Y68K mutant that substantially reduces prolyl-isomerase activity (*46*) and this too could rescue lysis. Finally, we purified a complex of protein E using the Epos rescue mutants (R3H and L19F) (*45*) in the Δ*slyD* strain. This YE complex was very unstable and aggregated according to SEC (fig. S2H). These results support the conclusion that the primary role of SlyD in promoting lysis is not prolyl-isomerization, but to protect protein E and stabilize the YES complex. This does not obviate a role for proline binding in complex assembly. Mutation of P65, in the FKBP active site, creates a slower-lysis phenotype (*22*), which may indicate that binding of these residues by SlyD is key to assembling the YES complex.

Protein E is evolutionarily constrained

Protein E arrived late in the evolution of ΦX174 and was overprinted into a +1 reading frame in the ORF for gene *D* (*47*, *48*). The structure supports that this embedding constrains the evolution of protein E (*12*). Considering the sequence changes across protein E isoforms, we note that from the N terminus through residue 70, with few exceptions, each position that is not completely conserved is either silent or that there is only a slight change in protein D (fig. S8A). The silent changes vary

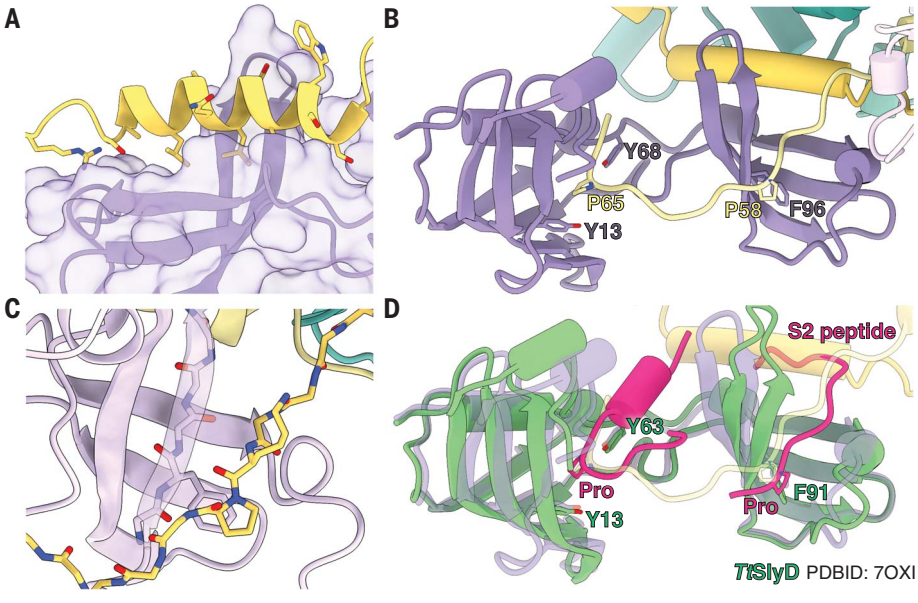


Fig. 5. Interactions between protein E and SlyD. (**A**) The amphipathic helix of protein E (yellow) as a illustration with side chains contacting SlyD shown as sticks. SlyD (purple) is shown as an illustration along with a transparent accessible surface (**B**) Full view of SlyD bound to two protein E molecules shown in different shades of yellow. (**C**) As in (A) highlighting the β-augmentation of the extended C-terminal protein E. (**D**) Structures of SlyD from the YES_{ID21} complex and *Tt*SlyD [PDBID:7OXI (*33*)] (green) aligned to the IF domains. The two S2 peptides bound to the *Tt*SlyD are shown in pink.

the codon's second position, which is the wobble position in the overlapping codon for protein D. For example, residue W7 in ΦX174 is replaced with a Ser or Leu in other species. Although seemingly substantial changes, each is coded by the sequence UXG and all three variations at this position are sampled. These variable positions generally do not contact MraY. An exception is found in the G4 isoform where a phenyalanine occurs at position 19, as in the Epos mutant, which results in a change from alanine to serine at position 79 in protein D. This places a polar residue in the hydrophobic core of protein D (fig. S8). This single nucleotide change increases protein E affinity to MraY but likely lowers the stability of protein D and the overall fitness of the virus. Introduction of this mutation results in smaller plaque formation relative to the wild type (*23*). The C terminus of protein D and the extensions in the longer isoforms of protein E are likely disordered and less constrained, hence the increased sequence variability (fig. S8).

Conclusions

The mechanism for ΦX174 inhibition of peptidoglycan biosynthesis is suprisingly simple. Protein E binds in the active site cleft preventing access of the lipid substrate and blocking conformational changes needed for catalysis (fig. S14). The YES complex structure provides a template for interpreting all the previous results for lysis by ΦX174. Notable among these are that the extensive functional mutations in protein E and MraY map to the interface between the two proteins. SlyD binds to the cytoplasmic domain of protein E to stabilize the MraY/protein E complex. As we demonstrated, the requirement for SlyD is based on the protein binding properties, as distantly related SlyD homologs can rescue lysis in a Δ*slyD* strain. SlyD in this case does not serve directly in the mechanism of inhibition of MraY and can be functionally replaced.

Our structures offer some insights for future work applying SGL proteins as therapies. Protein E binding to *Ec*MraY identifies a distinct mechanism from current inhibitors. Blocking of the lipid substrate access to the active site in MraY could be exploited by small molecules, with structure-based drug design guided by the protein E structure. Expanding protein E efficacy to Gram-positive bacteria and, potentially, other related polyisoprenyl-phosphotransferases such as bacterial WecA or the mammalian DPAGT1 could further increase applicability. SGL proteins, which have distinctly evolved many times (*49*), can also be used to improve the antibacterial potency of bacteriophages. The use of phages for medical therapies, although known for a hundred years, has attracted recent attention (*1*). Improving the efficiency of ΦX174 protein E based on structural information may help these therapies achieve their greatest potential in the fight against antimicrobial-resistant pathogens.

Materials and Methods

Coexpression of EcMraY, protein E, and EcSlyD

Δ*slyD* BL21(DE3)-competent cells were co-transformed with pET22b-SlyD$_{154}$ and either pRSFDuet*Ec*MraY-E$_{ID21}$ or pRSFDuet-*Ec*MraY-E$_{ΦX174}$ and plated in LB-agar containing 35 μg/ml Kanamycin and 100 μg/mL Ampicillin. Our pET22b-SlyD$_{154}$ construct expresses *E. coli* SlyD, modified by the removal of the flexible C-terminus. The pRSFDuet-*Ec*MraY-E$_{ID21}$ plasmid contains the ID21 isoform of protein E, along with a wild-type *Ec*MraY to prevent cell lysis from the overexpression of protein E. Cells were grown in 2xYT media at 37°C, 225 r.p.m., and induced at an OD$_{600}$ of 0.9 with 0.4mM IPTG at 18°C overnight. The culture was harvested by centrifugation for 10 minutes at 9000*xg*, 4°C then frozen or used immediately for purification.

Purification of the YES complex

The cells were resuspended in lysis buffer (20mM Tris-HCl pH 7.5, 300 mM NaCl, 10% Glycerol, 5mM β-mercaptoethanol (βME), 0.1mM PMSF, 0.1mM Benzamidine) and homogenized using a M-110L microfluidizer (Microfluidics). The lysate was cleared by a 20-minute centrifugation at a speed of 22,000*xg*. The supernatant was then centrifugated at 167,424*xg* and the resulting membrane pellet was then solubilized in the extraction buffer [10 mM HEPES pH 7.5, 300 mM NaCl, 5% Glycerol, 5mM βME, 0.1mM phenylmethylsulfonyl fluoride (PMSF), 0.1mM benzamidine, 10 mM imidazole and 1% dodecyl 4-O-α-D-glucopyranosyl-β-D-glucopyranoside (DDM)] After allowing for extraction for 1.5 hours at 4°C, the solution was centrifuged at 167,424*xg* for 30 minutes and the remaining lysate was mixed with 1mL NiNTA resin (Qiagen, Alameda, CA) then nutated at 4°C for two hours. This solution was loaded onto a gravity column and then washed with five column volumes of wash buffer (10 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 5mM βME, & 0.03% DDM) with 10mM imidazole followed by five column volumes of wash buffer with 30 mM imidazole. The YES complex was eluted in 20mL of wash buffer containing 200 mM imidazole. The final purification step was SEC (Superdex 200 10/300 GL, Millipore Sigma) in 10mM HEPES pH 7.5, 75 mM NaCl, 5% Glycerol, 5mM βME and 0.03% DDM. Fractions were assessed by SDS-PAGE and directly used for cryo-EM sample preparation.

Co-expression of EcMraY and protein E in various SlyD backgrounds

The pRSFDuet-E-*Ec*MraY and pRSFDuet-Epos-*Ec*MraY expression vectors were transformed into BL21-Star cells (Novagen). Similarly, the pRSFDuet-E(C-term)-SlyD$_{154}$ was transformed into SlyD-knockout cells. The cultures were grown at 37 °C to an OD$_{600}$ 0.8 and induced with 1 mM IPTG. Induced cultures were grown for 3 hours followed by harvesting by centrifugation at 9,000*xg* for 20 min. Cell pellets were resuspended in lysis buffer and lysed by sonication. The lysate was then cleared by centrifugation at 22,000*xg*, followed by a second centrifugation at 234,78*xg* for 1 hour to isolate the membrane fraction. The complex was extracted in 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 10% Glycerol, 10 mM Imidazole, and 1% n-Decyl-β-Maltoside (DM) and incubated at 4°C for 1.5 hours. The debris was cleared by centrifugation at 234,788*xg* for 30 min. The sample was incubated with 1 mL NiNTA resin for 1 hour, followed by a wash with 50 column volumes lysis buffer with 30mM Imidazole. The protein E complexes were similarly eluted in 300mM Imidazole. The elutions were concentrated and further purified by size exclusion chromatography (Superdex 200 10/300 GL, Millipore Sigma).

Lysis assays of WT protein E ΦX174 and ID21

LEMO DE3 competent cells were transformed with a pRSF-Duet vector either empty, with protein E$_{ΦX174}$, or protein E$_{ID21}$. Cultures were grown to an OD$_{600}$ of 0.2 and inoculated into a Corning 96-well Clear Flat Bottom plates in 100μL triplicate aliquots and induced as described previously. Cultures were incubated at 37C with orbital shaking at 220rpm using an Infinite M Nano+ (Tecan, Switzerland). Readings were taken in 5-minute intervals for 90 minutes.

Lysis assay for protein E constructs

LEMO DE3 competent cells (New England Biolabs, MA, USA) were transformed with a pRSFDuet vector either empty, with C-terminally FLAG tagged protein E$_{ΦX174}$ variants (WT, P21A, K46A). The lysis assays were performed in three biological replicates as previously described (*21*). Absorbance readings were recorded in 5-minute intervals for 1 hour and 30 minutes. Manual readings were taken using a Biowave Cell Density Meter CO8000. The values were plotted using GraphPad Prism version 9.1.1 for macOS.

Lysis assays based on SlyD variants

Δ*slyD* (*18*) cells were transformed with either a control empty pRSF-Duet vector or pRSFDuet-Protein$_{ΦX174}$ and either pET22b-*Ec*SlyD, pET22b-SlyD$_{154}$, pET22b-*Ec*SlyD Y68K, or pET22b-*Thermus thermophilus* SlyD. Cultures were grown in 2xYT media at 37°C and induced with 0.4mM IPTG once at an OD$_{600}$ of 0.2. Absorbance measurements were manually recorded in 5-minute intervals for 70 minutes. Similarly, Δ*slyD* cells were transformed with either a control empty pRSF-Duet vector or pRSF-Duet-Protein$_{ID21}$ either alone, with pET22b*Ec*SlyD, or with pET22b-*Ec*SlyD$_{154}$ and induced with 0.4mM IPTG. Readings were recorded using an Infinite M Nano+ plate reader as described above.

Sample preparation for cryoEM

The YES complex was diluted to 5.0 mg/mL in 10 mM HEPES pH 7.5, 75 mM NaCl, 5% Glycerol, 5mM βME and 0.03% DDM. Additionally, the YES_{ID21} sample was supplemented with 1mM *E. coli* total lipid extract (Avanti Polar Lipids, 100600P). Quantifoil holey carbon films R1.2/1.3 300 Mesh, Copper (Quantifoil, Micro Tools GmbH) grids were glow discharged with a 2-minute 20Å plasma current using a Pelco easiGlow, Emitech K100X. Grids were prepared using a Vitrobot (FEI Vitrobot Mark v4 x2, Mark v3) by applying 3μL of sample onto the grid followed by a 3.5 second blot using a +8-blot force and plunge frozen into liquid ethane.

Data acquisition and analysis

The grids were imaged in a 300 kV cryo-TEM microscope equipped with a Gatan K3 6k x 4k direct electron detector and a Gatan Energy Filter (slit width 20eV) in super-resolution mode using Serial EM. Datasets were collected at a 105k magnification with a pixel size of 0.416 Å/pixel. Movies with 40 frames were recorded with a total exposure dose of 60 $e^-/Å^2$ and a defocus range of -1.0 to -2.5 μm. For the YES_{ID21} complex, a total of 12,070 movies were recorded. Movies were normalized by gain reference and motion corrected using the patch motion correction built in function in cryosparc (v3.3.2) with a twofold bin that resulted in a pixel size of 0.832 Å/pixel (*50*). The contrast transfer function (CTF) was estimated using CTFFIND4 (*51*). Micrographs were manually curated, and low-quality images were removed for further analyses. A total of 2,462,335 particles were obtained followed by the generation of 6 ab-initio models. Out of the 6 models, two models are selected for classification into "good" and "trash" volumes. All particles were then sorted in these two volumes through heterogeneous refinement using particles extracted with a 4x bin, which produced 6,589,696 good particles. Heterogeneous refinement was used in an iterative manner to sort the particles into the 5 volumes (4 good and 1 trash). The 1,151,777 good particles were used for non-uniform homogeneous refinement to generate a higher resolution volume. The particles were then extracted with a 3x bin and sorted into 4 iterations of the higher resolution volume and 1 trash volume. Iterative rounds of heterogeneous refinement at 3x bin produced 935,754 particles. Particles were then extracted in a 2x bin and heterogeneously refined into either high- or low-resolution volumes. At this point, discerning features in the soluble region of the model were used to select the most complete volumes. The volumes were individually refined through non-uniform refinement and the particles that composed the volumes with most complete and highest resolution were used. A total of 122,452 particles were used for the most complete model obtained upon non-uniform refinement. The FSC-masked resolution was 3.5Å, while the unmasked resolution was 3.9Å. The half-maps were then used for post-processing through DeepEMhancer (*52*) with the high-resolution model selected for our most complete density map. Post-processing through DeepEMhancer removed the micelle and improved the features on the soluble portions of the map, however the lipid densities were also removed. The lipid densities described in this work are those of the YES_{ID21} map before post-processing. Notably, the dimer-interface lipid density was also present in the $YES_{ΦX174}$ density map without the supplementation of *E. coli* lipid extract. Figure 1D uses the densities before post-processing for the MraY dimer, micelle and lipid densities, and the DeepEMhancer post-processed map for protein E and SlyD. Supplemental figures S9 and S12 were made with the map before DeepEMhancer sharpening. The $YES_{ΦX174}$ complex dataset was processed in this same manner. A total of 10,798 movies were recorded. The model for the YES_{ID21} complex was then used as a template for template picking, from which 1,516,368 particles were picked and curated. Following gradual un-binning and sorting into good and trash volumes, 155,270 particles were used for the final iteration of non-uniform refinement. The local resolution of both maps was performed on cryosparc (v3.3.2). The half-maps were then post-processed using DeepEMhancer as described previously.

Model building

For starting models we used the *Aquifex aeolicus* un-bound structure (PDBID:4J72) (*25*) for *Ec*MraY and the *E. coli* NMR structure (PDBID:2K8I) (*31*) for *Ec*SlyD which were fit using phenix.dock. SlyD was then split into its two domains, IF and FKBP, at residues Y68 and G127. Protein E was modeled de novo up to residue P65 using Coot 0.8.9.2. The structures of the *Ec*MraY, protein E, and *Ec*SlyD-IF domain were refined using phenix.real space refinement and ISOLDE 1.6, and validated with PHENIX-1.19.2. After the refinements of *Ec*MraY, protein E, and the *Ec*SlyD-IF domain were completed, the FKBP domains were docked into density using ChimeraX. The complete YES complex structure was then refined using PHENIX-1.19.2 and ISOLDE 1.6. RMSDs were calculated using ChimeraX Matchmaker chain alignment. Structure figures were made using ChimeraX and sequence alignments using Jalview and ClustalW (*53–55*).

REFERENCES AND NOTES

1. S. A. Strathdee, G. F. Hatfull, V. K. Mutalik, R. T. Schooley, Phage therapy: From biological mechanisms to future directions. *Cell* **186**, 17–31 (2023). doi: 10.1016/j.cell.2022.11.017; pmid: 36608652
2. R. T. Schooley *et al*., Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant *Acinetobacter baumannii* Infection. *Antimicrob. Agents Chemother.* **61**, e00954–e17 (2017). doi: 10.1128/AAC.00954-17; pmid: 28807909
3. C. Ferriol-González, P. Domingo-Calap, Phage Therapy in Livestock and Companion Animals. *Antibiotics* **10**, 559 (2021). doi: 10.3390/antibiotics10050559; pmid: 34064754
4. K. E. Kortright, B. K. Chan, J. L. Koff, P. E. Turner, Phage therapy: A renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **25**, 219–232 (2019). doi: 10.1016/j.chom.2019.01.014; pmid: 30763536
5. F. D. Herelle, [The bacteriophage]. *Atomes* **3**, 399–403 (1948). pmid: 18104413
6. R. L. Sinsheimer, A single-stranded deoxyribonucleic acid from bacteriophage ΦX174. *J. Mol. Biol.* **1**, 43–IN6 (1959). doi: 10.1016/S0022-2836(59)80006-1
7. C. A. Hutchison 3rd, R. L. Sinsheimer, The process of infection with bacteriophage phi-X174: X. Mutations in a phi-X Lysis gene. *J. Mol. Biol.* **18**, 429–447 (1966). doi: 10.1016/S0022-2836(66)80035-9; pmid: 5968177
8. R. M. Benbow, C. A. HutchisonIII, J. D. Fabricant, R. L. Sinsheimer, Genetic map of bacteriophage ΦX174. *J. Virol.* **7**, 549–558 (1971). doi: 10.1128/jvi.7.5.549-558.1971; pmid: 16789129
9. H. O. Smith, C. A. Hutchison3rd, C. Pfannkoch, J. C. Venter, Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15440–15445 (2003). doi: 10.1073/pnas.2237126100; pmid: 14657399
10. M.-S. Kim, E.-J. Park, S. W. Roh, J.-W. Bae, Diversity and abundance of single-stranded DNA viruses in human feces. *Appl. Environ. Microbiol.* **77**, 8062–8070 (2011). doi: 10.1128/AEM.06331-11; pmid: 21948823
11. F. Sanger *et al*., Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* **265**, 687–695 (1977). doi: 10.1038/265687a0; pmid: 870828
12. B. G. Barrell, G. M. Air, C. A. Hutchison 3rd, Overlapping genes in bacteriophage phiX174. *Nature* **264**, 34–41 (1976). doi: 10.1038/264034a0; pmid: 1004533
13. K. D. Young, R. Young, Lytic action of cloned phi X174 gene E. *J. Virol.* **44**, 993–1002 (1982). doi: 10.1128/jvi.44.3.993-1002.1982; pmid: 6294347
14. W. Lubitz, G. Halfmann, R. Plapp, Lysis of Escherichia coli after infection with phiX174 depends on the regulation of the cellular autolytic system. *J. Gen. Microbiol.* **130**, 1079–1087 (1984). doi: 10.1099/00221287-130-5-1079; pmid: 6236279
15. A. Witte, G. Wanner, M. Sulzner, W. Lubitz, Dynamics of PhiX174 protein E-mediated lysis of Escherichia coli. *Arch. Microbiol.* **157**, 381–388 (1992). doi: 10.1007/BF00248685; pmid: 1534215
16. T. G. Bernhardt, W. D. Roof, R. Young, Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4297–4302 (2000). doi: 10.1073/pnas.97.8.4297; pmid: 10760296
17. D. Maratea, K. Young, R. Young, Deletion and fusion analysis of the phage phi X174 lysis gene E. *Gene* **40**, 39–46 (1985). doi: 10.1016/0378-1119(85)90022-8; pmid: 2936651
18. W. D. Roof, S. M. Horne, K. D. Young, R. Young, slyD, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. *J. Biol. Chem.* **269**, 2902–2910 (1994). doi: 10.1016/S0021-9258(17)42027-8; pmid: 8300625
19. K. R. Chamakura, R. Young, Single-gene lysis in the metagenomic era. *Curr. Opin. Microbiol.* **56**, 109–117 (2020). doi: 10.1016/j.mib.2020.09.015; pmid: 33075663
20. K. J. Buckley, M. Hayashi, Lytic activity localized to membrane-spanning region of phi X174 E protein. *Mol. Gen. Genet.* **204**, 120–125 (1986). doi: 10.1007/BF00330198; pmid: 3018438
21. S. Tanaka, W. M. Clemons Jr., Minimal requirements for inhibition of MraY by lysis protein E from bacteriophage ΦX174. *Mol. Microbiol.* **85**, 975–985 (2012). doi: 10.1111/j.1365-2958.2012.08153.x; pmid: 22742425
22. A. Witte, G. Schrot, P. Schön, W. Lubitz, Proline 21, a residue within the α-helical domain of phiX174 lysis protein E, is required for its function in Escherichia coli. *Mol. Microbiol.* **26**, 337–346 (2003). doi: 10.1046/j.1365-2958.1997.5781941.x; pmid: 9383158
23. T. G. Bernhardt, W. D. Roof, R. Young, The Escherichia coli FKBP-type PPIase SlyD is required for the stabilization of the E lysis protein of bacteriophage φ X174. *Mol. Microbiol.* **45**, 99–108 (2002). doi: 10.1046/j.1365-2958.2002.02984.x; pmid: 12100551
24. A. C. Teo, D. I. Roper, Core steps of membrane-bound peptidoglycan biosynthesis: Recent advances, insight and opportunities. *Antibiotics* **4**, 495–520 (2015). doi: 10.3390/antibiotics4040495; pmid: 27025638
25. B. C. Chung *et al*., Crystal structure of MraY, an essential membrane enzyme for bacterial cell wall synthesis. *Science*

341, 1012–1016 (2013). doi: 10.1126/science.1236501; pmid: 23990562
26. E. H. Mashalidis *et al*., Chemical logic of MraY inhibition by antibacterial nucleoside natural products. *Nat. Commun.* **10**, 2917 (2019). doi: 10.1038/s41467-019-10957-9; pmid: 31266949
27. J. K. Hakulinen *et al*., MraY-antibiotic complex reveals details of tunicamycin mode of action. *Nat. Chem. Biol.* **13**, 265–267 (2017). doi: 10.1038/nchembio.2270; pmid: 28068312
28. B. Al-Dabbagh *et al*., Active site mapping of MraY, a member of the polyprenyl-phosphate N-acetylhexosamine 1-phosphate transferase superfamily, catalyzing the first membrane step of peptidoglycan biosynthesis. *Biochemistry* **47**, 8919–8928 (2008). doi: 10.1021/bi8006274; pmid: 18672909
29. V. V. Mokhonov *et al*., SlyD-deficient *Escherichia coli* strains: A highway to contaminant-free protein extraction. *Biochem. Biophys. Res. Commun.* **499**, 967–972 (2018). doi: 10.1016/j.bbrc.2018.04.029; pmid: 29626483
30. C. Löw *et al*., Crystal structure determination and functional characterization of the metallochaperone SlyD from *Thermus thermophilus*. *J. Mol. Biol.* **398**, 375–390 (2010). doi: 10.1016/j.jmb.2010.03.014; pmid: 20230833
31. U. Weininger *et al*., NMR solution structure of SlyD from Escherichia coli: Spatial separation of prolyl isomerase and chaperone function. *J. Mol. Biol.* **387**, 295–305 (2009). doi: 10.1016/j.jmb.2009.01.034; pmid: 19356587
32. E. M. Quistgaard *et al*., Molecular insights into substrate recognition and catalytic mechanism of the chaperone and FKBP peptidyl-prolyl isomerase SlyD. *BMC Biol.* **14**, 82 (2016). doi: 10.1186/s12915-016-0300-3; pmid: 27664121
33. S. Pazicky, A. A. Werle, J. Lei, C. Löw, U. Weininger, Impact of distant peptide substrate residues on enzymatic activity of SlyD. *Cell. Mol. Life Sci.* **79**, 138 (2022). doi: 10.1007/s00018-022-04179-4; pmid: 35184231
34. L. Martino *et al*., The interaction of the *Escherichia coli* protein SlyD with nickel ions illuminates the mechanism of regulation of its peptidyl-prolyl isomerase activity. *FEBS J.* **276**, 4529–4544 (2009). doi: 10.1111/j.1742-4658.2009.07159.x; pmid: 19645725
35. Y. Zheng, D. K. Struck, T. G. Bernhardt, R. Young, Genetic analysis of MraY inhibition by the phiX174 protein E. *Genetics* **180**, 1459–1466 (2008). doi: 10.1534/genetics.108.093443; pmid: 18791230
36. Y. Zheng, D. K. Struck, R. Young, Purification and functional characterization of phiX174 lysis protein E. *Biochemistry* **48**, 4999–5006 (2009). doi: 10.1021/bi900469g; pmid: 19379010
37. B. C. Chung *et al*., Structural insights into inhibition of lipid I production in bacterial cell wall synthesis. *Nature* **533**, 557–560 (2016). doi: 10.1038/nature17636; pmid: 27088606
38. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
39. P. Kudela, V. J. Koller, W. Lubitz, Bacterial ghosts (BGs)—Advanced antigen and drug delivery system. *Vaccine* **28**, 5760–5767 (2010). doi: 10.1016/j.vaccine.2010.06.087; pmid: 20619379
40. U. B. Mayr *et al*., Bacterial ghosts as antigen delivery vehicles. *Adv. Drug Deliv. Rev.* **57**, 1381–1391 (2005). doi: 10.1016/j.addr.2005.01.027; pmid: 15878634
41. G. Halfmann, F. Götz, W. Lubitz, Expression of bacteriophage PhiX174 lysis gene E in Staphylococcus carnosus TM300. *FEMS Microbiol. Lett.* **108**, 139–143 (1993). doi: 10.1111/j.1574-6968.1993.tb06089.x; pmid: 8486239
42. E. Henrich *et al*., Lipid Requirements for the Enzymatic Activity of MraY Translocases and in Vitro Reconstitution of the Lipid II Synthesis Pathway. *J. Biol. Chem.* **291**, 2535–2546 (2016). doi: 10.1074/jbc.M115.664292; pmid: 26620564
43. E. Henrich *et al*., Analyzing native membrane protein assembly in nanodiscs by combined non-covalent mass spectrometry and synthetic biology. *eLife* **6**, e20954 (2017). doi: 10.7554/eLife.20954; pmid: 28067619
44. A. O. Oluwole *et al*., Peptidoglycan biosynthesis is driven by lipid transfer along enzyme-substrate affinity gradients. *Nat. Commun.* **13**, 2278 (2022). doi: 10.1038/s41467-022-29836-x; pmid: 35477938
45. W. D. Roof, R. Young, Φ X174 lysis requires slyD, a host gene which is related to the FKBP family of peptidyl-prolyl cis-trans isomerases. *FEMS Microbiol. Rev.* **17**, 213–218 (1995). doi: 10.1016/0168-6445(94)00073-5; pmid: 7669348
46. T. Ikura, N. Ito, Requirements for peptidyl-prolyl isomerization activity: A comprehensive mutational analysis of the substrate-binding cavity of FK506-binding protein 12. *Protein Sci.* **16**, 2618–2625 (2007). doi: 10.1110/ps.073203707; pmid: 18029417
47. J. C. Fiddes, G. N. Godson, Evolution of the three overlapping gene systems in G4 and phi X174. *J. Mol. Biol.* **133**, 19–43 (1979). doi: 10.1016/0022-2836(79)90249-3; pmid: 160951
48. A. Pavesi, Origin, evolution and stability of overlapping genes in viruses: A systematic review. *Genes* **12**, 809 (2021). doi: 10.3390/genes12060809; pmid: 34073395
49. K. R. Chamakura *et al*., Rapid de novo evolution of lysis genes in single-stranded RNA phages. *Nat. Commun.* **11**, 6009 (2020). doi: 10.1038/s41467-020-19860-0; pmid: 33243984
50. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017). doi: 10.1038/nmeth.4169; pmid: 28165473
51. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015). doi: 10.1016/j.jsb.2015.08.008; pmid: 26278980
52. R. Sanchez-Garcia *et al*., DeepEMhancer: A deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021). doi: 10.1038/s42003-021-02399-1; pmid: 34267316
53. E. F. Pettersen *et al*., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021). doi: 10.1002/pro.3943; pmid: 32881101
54. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009). doi: 10.1093/bioinformatics/btp033; pmid: 19151095
55. M. Steinegger, J. Söding, Clustering huge protein sequence sets in linear time. *Nat. Commun.* **9**, 2542 (2018). doi: 10.1038/s41467-018-04964-5; pmid: 29959318

ACKNOWLEDGMENTS

We are particularly grateful to R. Young and T. Bernhardt for providing inspiration and feedback during the course of this project. We thank them, M. Kurosu, and D. Rees for critical feedback on the manuscript. We further thank R. Young for providing the Δ*slyD* strain. (Cryo) Electron microscopy was done in the Beckman Institute Resource Center for Transmission Electron Microscopy at Caltech. We are grateful to S. Chen for help with data collection and processing. **Funding:** This work was funded by the following: National Institutes of Health grant R01GM114611 (to W.M.C. and M.K.); National Institutes of Health grant DP1GM105385 (to W.M.C.); and the G. Harold and Leila Y. Mathers Foundation (to W.M.C.) **Author contributions:** Conceptualization: S.T. and W.M.C. Methodology: A.K.O., N.R., Y.E.L., S.T., H.G.Y., L.K., and W.M.C. Investigation: A.K.O., N.R., Y.E.L., S.T., H.G.Y., L.K., W.M.C. Visualization: A.K.O., N.R., S.T., and W.M.C. Funding acquisition: W.M.C. Project administration: W.M.C. Supervision: W.M.C. Writing – original draft: A.K.O., N.R., and W.M.C. Writing – review and editing: A.K.O., N.R., Y.E.L., S.T., H.G.Y., L.K., and W.M.C. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** All experimental data are available in the main text or the supplementary materials. Coordinates with experimental maps have been deposited to the RCSB or the EMDB for both the YES_{ID21} (PDB ID 8G01, EMDB-29641) and $YES_{\Phi X174}$ (PDB ID 8G02 and EMDB-29642) complexes. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg9091
Figs. S1 to S14
Table S1
References (*56*–*57*)
MDAR Reproducibility Checklist
Movie S1

View/request a protocol for this paper from *Bio-protocol*.

Submitted 4 February 2023; accepted 23 May 2023
10.1126/science.adg9091

RESEARCH ARTICLE

TOPOLOGICAL MATTER

Quantum metric nonlinear Hall effect in a topological antiferromagnetic heterostructure

Anyuan Gao[1], Yu-Fei Liu[1,2], Jian-Xiang Qiu[1], Barun Ghosh[3], Thaís V. Trevisan[4,5], Yugo Onishi[6], Chaowei Hu[7], Tiema Qian[7], Hung-Ju Tien[8], Shao-Wen Chen[2], Mengqi Huang[9], Damien Bérubé[1], Houchen Li[1], Christian Tzschaschel[1], Thao Dinh[1,2], Zhe Sun[1,10], Sheng-Chin Ho[1], Shang-Wei Lien[8], Bahadur Singh[11], Kenji Watanabe[12], Takashi Taniguchi[12], David C. Bell[13,14], Hsin Lin[15], Tay-Rong Chang[8,16,17], Chunhui Rita Du[9], Arun Bansil[3], Liang Fu[6], Ni Ni[7], Peter P. Orth[4,5], Qiong Ma[10,18], Su-Yang Xu[1]*

Quantum geometry in condensed-matter physics has two components: the real part quantum metric and the imaginary part Berry curvature. Whereas the effects of Berry curvature have been observed through phenomena such as the quantum Hall effect in two-dimensional electron gases and the anomalous Hall effect (AHE) in ferromagnets, the quantum metric has rarely been explored. Here, we report a nonlinear Hall effect induced by the quantum metric dipole by interfacing even-layered $MnBi_2Te_4$ with black phosphorus. The quantum metric nonlinear Hall effect switches direction upon reversing the antiferromagnetic (AFM) spins and exhibits distinct scaling that is independent of the scattering time. Our results open the door to discovering quantum metric responses predicted theoretically and pave the way for applications that bridge nonlinear electronics with AFM spintronics.

Nonlinearities are crucial in many branches of physics, ranging from atomic physics to condensed-matter and complex dynamical systems. Nonlinear electrical transport is the foundation of applications such as rectification and wave mixing. Classically, the most well-known nonlinear device is a PN diode (Fig. 1A). Noncentrosymmetric polar materials (Fig. 1B) are similar to PN diodes as they both possess an electric dipole. They have recently been discovered to show intrinsic nonlinear electrical transport, which may not only lead to new nonlinear applications but also provide a powerful probe of the quantum geometry of the conduction electrons (*1*–*16*). Broadly, the nonlinear transport in both diodes (Fig. 1A) and noncentrosymmetric conductors (Fig. 1B) arises from an inversion asymmetric charge distributions (e.g., an electric dipole). Because the electron has another fundamental degree of freedom, spin, an interesting question is whether spin can also lead to an electrical nonlinearity even in a centrosymmetric lattice. One ideal platform is the class of parity-inversion time-reversal (*PT*)–symmetric antiferromagnets (AFMs) (*17*), where only the spins feature a noncentrosymmetric distribution (Fig. 1C).

Important clues can be drawn from previous optical experiments, in which optical second-harmonic generation (SHG) has been observed in the *PT*-symmetric AFMs, including Cr_2O_3 and CrI_3 (*18*). Nevertheless, nonlinear transport is distinct because it directly probes the Fermi surface electrons and in many cases their geometrical properties (*1*, *2*). As such, it enables a probe of the quantum geometry (*1*, *2*) of the topological bands at the Fermi level.

The quantum geometry has two parts, $T = g\text{-}i/2\Omega$ (*1*) (T is the quantum metric tensor). The imaginary part is the well-known Berry curvature $\Omega_{\alpha\beta} = -2\mathrm{Im}\sum_{m\neq n}\left[\langle u_n|i\partial_{k_\alpha}u_m\rangle\langle u_m|i\partial_{k_\beta}u_n\rangle\right]$, which describes the curvature of wave function in Hilbert space (n, m are band indices and α, β are spatial directions). Berry curvature has been identified as the source of many unusual electronic and optical responses. By contrast, the real part is the quantum metric, $g_{\alpha\beta} = \mathrm{Re}\sum_{m\neq n}\left[\langle u_n|i\partial_{k_\alpha}u_m\rangle\langle u_m|i\partial_{k_\beta}u_n\rangle\right]$, which measures the distance between neighboring Bloch wave functions in Hilbert space [i.e., see section IV.1 of (*19*)]. Although equally important, the quantum metric is much less explored. There have been a few examples, including the orbital magnetic susceptibilities (*20*), a third-order Hall effect (*13*), and the quantum metric in atomic physics (*21*). However, the way quantum metric regulates electronic motion remains largely unknown. Recently, theories have predicted a wide range of exotic quantum metric responses (*20*, *22*–*34*).

Here, we report the observation of the quantum metric dipole–induced second-order anomalous Hall effect (AHE) (*20*, *22*–*25*). In past decades, there have been numerous studies of the AHE (both linear and nonlinear) induced by Berry curvature. Recent theoretical studies, however, predict that the quantum metric can also lead to AHE, thereby advancing our understanding of the fundamental origin of the AHE. Distinct from the Berry curvature–induced AHEs, this effect is predicted to exist in the *PT*-symmetric AFMs (*20*, *22*–*25*), where *PT* forces the Berry curvature to vanish identically but the effects of quantum metric can manifest strongly. We design and fabricate a feasible material platform and demonstrate the appearance of the quantum metric nonlinear Hall effect (NHE). To conceptualize this type of NHE, we draw comparison with the well-known AHE in ferromagnetic metals (*35*), where Berry curvature leads to the anomalous velocity and therefore the AHE, $v_{\mathrm{anomalous}} \propto \int_{\mathbf{k}} \mathbf{E}_{||} \times \boldsymbol{\Omega}$ ($\mathbf{E}_{||}$ is the in-plane source-drain electric field). By contrast, in a *PT*-symmetric AFM, Berry curvature is zero as dictated by *PT*. However, a nonzero quantum metric g can induce an anomalous velocity to second order in $\mathbf{E}_{||}$, $v_{\mathrm{anomalous}} \propto \int_{\mathbf{k}} \mathbf{E}_{||} \times \left[\nabla_{\boldsymbol{k}} \times \left(g\mathbf{E}_{||}\right)\right]$, as proposed in (*20*). This leads to an intrinsic second-order Hall effect. From the expression above, one can show that this effect is nonzero only when the system breaks both P and T. Therefore, we need *PT*-symmetric AFM conductors with a large quantum metric on the Fermi surface. We have carefully considered possible materials, and identified two-dimensional (2D) even-layered $MnBi_2Te_4$ (*15*, *36*–*46*) as an ideal platform. Even-layered $MnBi_2Te_4$ is a *PT*-symmetric AFM. Moreover, its topological bands support gate-tunable transport and a giant quantum metric (*19*). However, its lattice has C_{3z} rotational symmetry (Fig. 1, D and E), which forces the effect to vanish (*22*). To break C_{3z}, we interface $MnBi_2Te_4$ with black phosphorus (BP) (*47*).

Demonstration of rotational symmetry breaking

We start by showing that interfacing $MnBi_2Te_4$ with BP indeed breaks its C_{3z} rotational symmetry. To this end, we study the directional dependence of the resistance (*6*) of $MnBi_2Te_4$ without and with BP. We fabricated a six-septuple-layer

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [2]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [3]Department of Physics, Northeastern University, Boston, MA 02115, USA. [4]Department of Physics and Astronomy, Iowa State University, Ames, IA 50011, USA. [5]Ames National Laboratory, Ames, IA 50011, USA. [6]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Department of Physics and Astronomy and California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA. [8]Department of Physics, National Cheng Kung University, Tainan 701, Taiwan. [9]Department of Physics, University of California San Diego, La Jolla, CA, 92093, USA. [10]Department of Physics, Boston College, Chestnut Hill, MA, USA. [11]Department of Condensed Matter Physics and Materials Science, Tata Institute of Fundamental Research, Colaba, Mumbai, India. [12]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [13]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [14]Center for Nanoscale Systems, Harvard University, Cambridge, MA 02138, USA. [15]Institute of Physics, Academia Sinica, Taipei 11529, Taiwan. [16]Center for Quantum Frontiers of Research and Technology (QFort), Tainan 701, Taiwan. [17]Physics Division National Center for Theoretical Sciences, Taipei 10617, Taiwan. [18]Canadian Institute for Advanced Research, Toronto, Canada.
*Corresponding author. Email: suyangxu@fas.harvard.edu

(6SL) $MnBi_2Te_4$ device with radially distributed electrical contacts (Device-BM1). The four-probe resistance (T = 1.8 K) is found to be isotropic (Fig. 1G, blue curve), consistent with the C_{3z} symmetry. We then stacked a BP layer (~10 nm) onto this $MnBi_2Te_4$ sample and performed the measurements again. The resistance develops a clear anisotropy with a 180° periodicity (Fig. 1G, red curve), providing a clear signature of the breaking of C_{3z} [section I.4 of (*19*) shows that the transport is dominated by the $MnBi_2Te_4$ layer of the heterostructure]. The transverse resistance and two-probe resistance also show the breaking of C_{3z} [fig. S12 of (*19*)]. We further substantiate the breaking of C_{3z} symmetry by an independent method, the optical SHG, at room temperature. As shown in Fig. 1H, our SHG data also show the clear breaking of C_{3z} symmetry [see detailed discussions in section I.5 of (*19*)]. Our demonstration of C_{3z} breaking establishes the BP/$MnBi_2Te_4$ heterostructure as an ideal platform to search for this effect.

Observation of the nonlinear Hall effect

To measure the linear and nonlinear electrical transport, we pass a current at frequency ω (I^{ω}) and use the lock-in technique to detect linear voltage V^{ω} and nonlinear voltage $V^{2\omega}$. We describe the nonlinear voltage as $V^{2\omega}_{ijk}$, where i is the direction of the nonlinear voltage $V^{2\omega}$ and j, k are the directions of the injected current I^{ω}. All measurements are performed at B = 0.

Figure 1I shows the nonlinear Hall voltage $V^{2\omega}_{yxx}$ of the Device-BM1 before and after being interfaced with BP. A prominent nonlinear Hall signal emerges only after BP is introduced. This is in sharp contrast to the linear voltage (Fig. 1I, inset), which becomes even slightly smaller upon the introduction of BP. Such observation agrees well with the theoretical expectation of the intrinsic NHE induced by a quantum metric dipole. To exclude that the effect is caused by a Berry curvature dipole (*4*, *6*, *7*, *9*), we study the relationship between the second-order NHE and the AFM order in $MnBi_2Te_4$.

The AFM spin-induced nonlinearity

Overall, we have fabricated 30 BP/even-layered $MnBi_2Te_4$ heterostructure devices [see section I.0 of (*19*) for our systematic data that confirm the $MnBi_2Te_4$ thickness in our devices]. In all of the 30 devices, we have observed the NHE with consistent behaviors as a function of AFM order, spatial direction, scattering time, vertical electric field, and doping [see fig. S22 and table S1 for a summary of all 30 devices (*19*)]. Here, we focus on the Device-BMB1 (Fig. 2A), which has two-layer BP on both sides of six-septuple-layer $MnBi_2Te_4$. Moreover, we have ensured that the crystalline a axes of the BPs and the $MnBi_2Te_4$ are aligned (Fig. 2A). Such a carefully controlled configuration is important to preserve $MnBi_2Te_4$'s PT symmetry, which ensures that the Berry curvature and Berry curvature dipole vanish. Figure 2B shows a large transverse nonlinear response $V^{2\omega}_{yxx}$. We have also measured the longitudinal nonlinear response $V^{2\omega}_{xxx}$, which shows no observable signal. Therefore, our data reveal an interesting "Hall dominance" in the nonlinear transport.

We now focus on exploring how the nonlinear Hall signal depends on opposite AFM states. In ferromagnets, the opposite ferromagnetic states can be controlled by sweeping the B field. In PT-symmetric AFMs including Cr_2O_3, even-layered CrI_3, and even-layered $MnBi_2Te_4$ (*44*, *48*, *49*), previous works have shown that the opposite AFM states can be controlled by sweeping the vertical B_z field under a fixed vertical E_z field. Hence, we follow the previously established procedures (*44*): Under a fixed E_z (E_z = −0.17 V/nm), we sweep B_z from −8 T to 0 T or from +8 T to 0 T to prepare the two AFM states (Fig. 2, C and D). We first study the AFM I. The linear voltage V^{ω}_{xx} (Fig. 2E) exhibits a typical ohmic behavior. The nonlinear voltage $V^{2\omega}_{yxx}$ (Fig. 2G) is prominent and its sign is positive. We then prepare AFM II. The linear voltage V^{ω}_{xx} (Fig. 2F) remains unchanged. In sharp contrast, the nonlinear voltage $V^{2\omega}_{yxx}$ (Fig. 2H) flips sign. For both AFM I and II, the NHE is only present in the AFM phase (Fig. 2, I and J). Our observation that the nonlinear Hall signal flips sign upon reversing the AFM order further demonstrates its quantum metric dipole origin, because the quantum metric dipole is theoretically expected to be opposite for the opposite AFM domains [see section III of (*19*)]. Our nonlinear Hall signal measures an average over all AFM domains. However, our experiments suggest that our sample is prepared into predominantly one domain. If our sample consisted of opposite domains with a 50%-50% composition, then the measured nonlinear Hall signal would average to zero. By contrast,

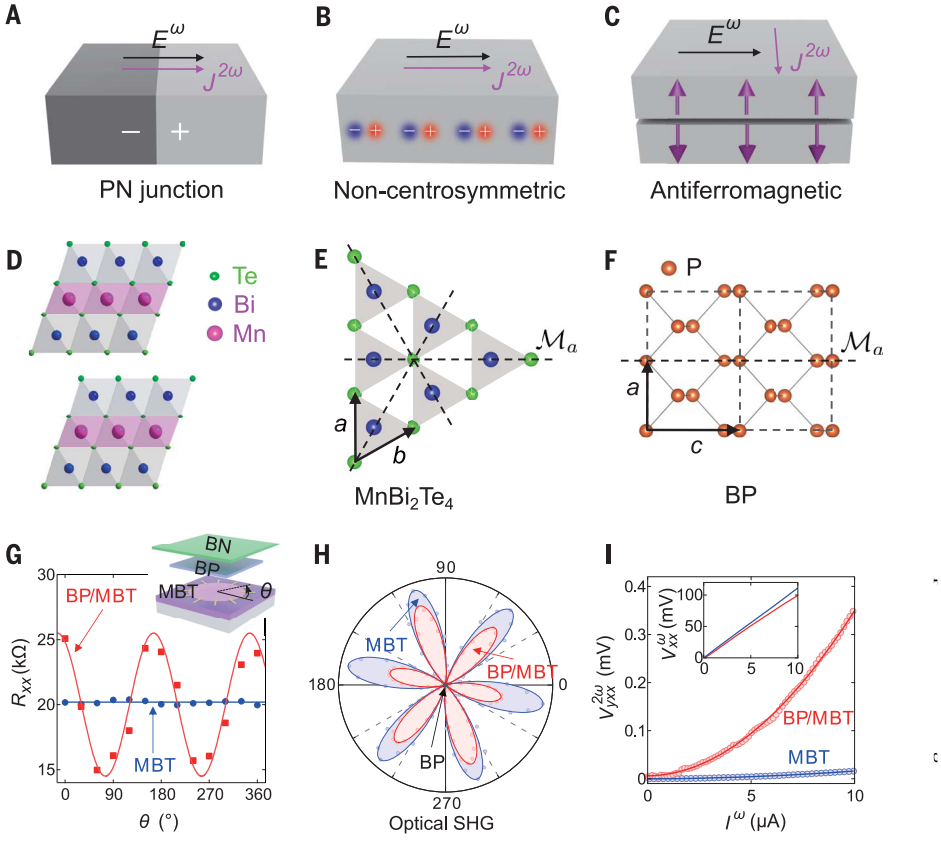


Fig. 1. Spin-induced electrical nonlinearity in *PT*-symmetric antiferromagnets and introduction to our sample. (**A** and **B**) Nonlinear electrical transport in PN junctions and noncentrosymmetric conductors (charge-induced electrical nonlinearity). (**C**) Nonlinear electrical transport in *PT*-symmetric AFMs (spin-induced electrical nonlinearity). (**D** to **F**) Lattice structures of the $MnBi_2Te_4$ and BP. (**G** and **H**) Angle-resolved resistance and optical SHG measurements of a 6SL $MnBi_2Te_4$ before and after being interfaced with BP. (**I**) The nonlinear Hall signal $V^{2\omega}_{yxx}$ before and after being interfaced with BP at B = 0 T. Inset: The linear longitudinal voltage V^{ω}_{xx} before and after being interfaced with BP. All data in Figs. 1 to 4 were taken at temperature T = 1.8 K unless otherwise noted. Data in Fig. 1 and Fig. 2B are from Device-BM1. Data in Fig. 2, E to J, Fig. 3, and Fig. 4A are from Device-BMB1. Data in Fig. 4F are from Device-BM21.

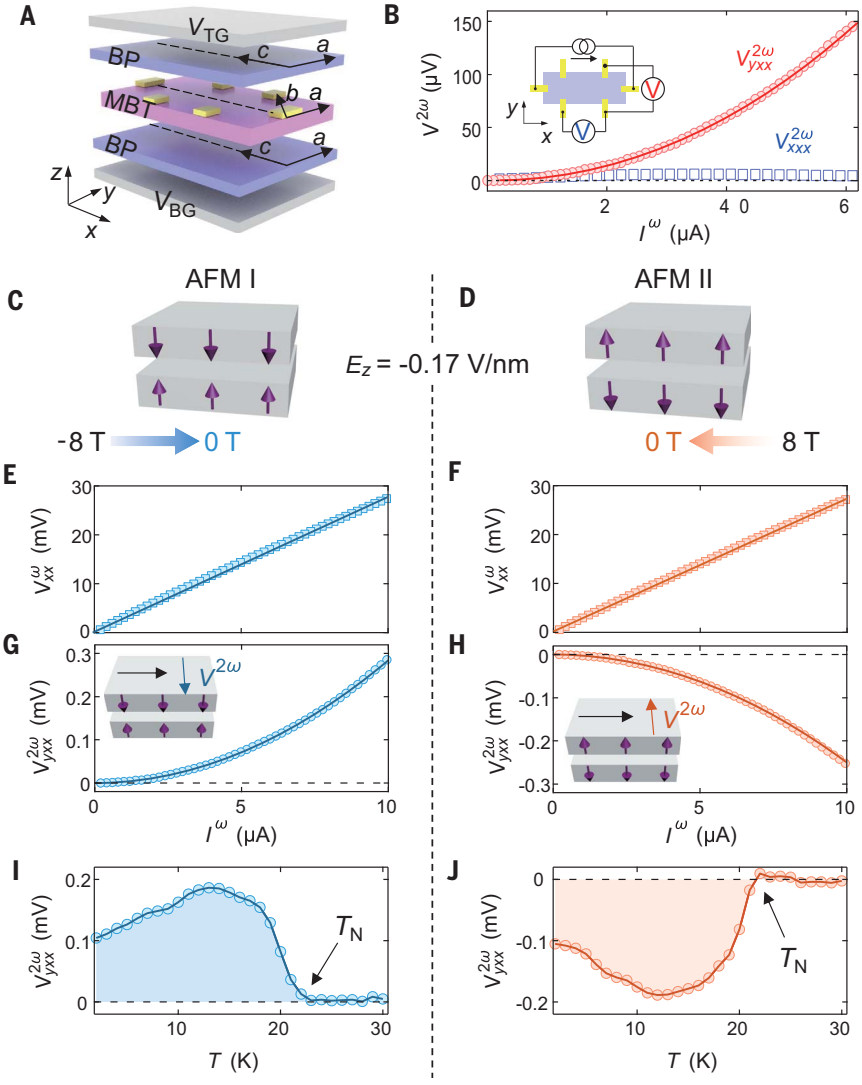


Fig. 2. Observation of the antiferromagnetic NHE. (**A**) Schematic illustration of 2L BP/6SL $MnBi_2Te_4$/2L BP device (Device-BMB1). The crystalline *a* axes of the BPs and the $MnBi_2Te_4$ were all aligned [fig. S16 of (*19*)]. (**B**) The longitudinal ($V^{2\omega}_{xxx}$) and Hall ($V^{2\omega}_{yxx}$) components of the nonlinear voltage. (**C** and **D**) We follow the procedures established by previous works (*44*): Under a fixed E_z (−0.17 V/nm), we sweep B_z from −8 T to 0 T or from +8 T to 0 T to prepare the two AFM states. (**E** and **F**) Linear longitudinal voltage as a function of incidence current for AFM I and AFM II. (**G** and **I**) Nonlinear Hall voltage as a function of incident current and temperature of AFM I, respectively. (**H** and **J**) The same as panels (G) and (I) but for AFM II.

our data show a large nonzero nonlinear Hall signal. Moreover, the sign of the observed signal flips as we prepare the opposite AFM domain. Further, the magnitude of the measured signal is consistent with the theoretically calculated value, which assumes a single domain. Spatially resolved magnetic measurements will be needed to determine the exact domain composition.

We now perform further systematic studies. Because the nonlinear Hall current flips sign upon reversing the AFM order, all the nonlinear Hall data (apart from Fig. 2) are obtained by taking the difference between the two AFM domains. First, the intrinsic NHE is expected to be independent of the scattering time, unlike other related responses. Similar to the intrinsic AHE in ferromagnetic metals (*35*), and in contrast to the quantum AHE, there is still dissipation through the linear Drude conductivity σ_{xx}, The nonlinear Hall conductivity can be directly extracted from our data by $\sigma^{2\omega}_{yxx} = J^{2\omega}_{yxx}/E^{\omega 2}_{x} = \left(V^{2\omega}_{yxx} l^3\right)/\left(I^{\omega 2}_{x} R^3_{xx} w^2 d\right)$, where l, w, d are the length, width, and thickness of the sample. Previous experiments have studied the scattering time τ dependence of various Hall effects (*6*, *9*, *14*, *35*) by investigating the scaling between the corresponding Hall conductivity and the Drude conductivity. Following this established method, we study the scaling between $\sigma^{2\omega}_{yxx}$ and σ_{xx}. Our data (Fig. 3A) show that $\sigma^{2\omega}_{yxx}$ is independent of σ_{xx} below ~15 K. Moreover, consistent results have been observed at multiple charge densities in the same sample and from different samples [section III.9 of (*19*)]. These systematic data point to the conclusion that the $\sigma^{2\omega}_{yxx}$ is independent of scattering time τ below ~15 K. Above ~15 K, $\sigma^{2\omega}_{yxx}$ vanishes quickly across Néel temperature T_N because the AFM order vanishes and our NHE only exists in the AFM phase. Hence, studying the τ dependence at temperatures near T_N would require one to take the strong influence of the AFM order near T_N into account [see section III.8 of (*19*) for additional measurements and analysis]. Second, the intrinsic NHE does not require a noncentrosymmetric lattice or any explicit breaking of *PT* symmetry. To test this, we explicitly break *PT* by applying a vertical E_z field via dual gating. The nonlinear Hall signal is already prominent even at $E_z = 0$ (Fig. 3D), confirming that it does not require any *PT* breaking. Moreover, the nonlinear Hall signal is symmetric for $\pm E_z$, also consistent with the expectation [see section IV.2 of (*19*)]. Third, the NHE is expected to be sensitive to the direction of the incident current I^{ω}. In Fig. 3B, we measure the nonlinear Hall conductivity as a function of the direction of I^{ω}. Indeed, the signal is most prominent when I^{ω} is along a particular in-plane direction. In this way, we experimentally mapped out the direction of the relevant geometrical dipole (the quantum metric dipole in our case, as we demonstrate next). Moreover, the intrinsic NHE is found to be independent of frequency [fig. S23 of (*19*)]. In principle, the frequency independence is expected to persist until ω is large enough to induce an interband transition (roughly terahertz or far-infrared). Future experiments are needed to test the NHE in that regime.

Excluding competing mechanisms

Although we tried to eliminate Berry curvature dipole by aligning the crystalline *a* axes between BPs and $MnBi_2Te_4$ to preserve *PT* symmetry (Fig. 2A), let us assume that the alignment is imperfect, so Berry curvature dipole is allowed. We now show that the observed relationship between the nonlinear Hall signal and AFM order can distinguish between Berry curvature dipole D_{Berry} and quantum metric dipole D_{Metric}. D_{Berry} can be understood as a distribution of the Berry curvature around the Fermi surface such that it is larger on one side of the Fermi surface than on the opposite side. A similar picture holds for D_{Metric} (Fig. 3). As we observe that the nonlinear Hall signal changes sign upon the reversal of AFM order, the dipole that causes our observed nonlinear Hall signal must also flip. Let us assume that the AFM I has $D_{\text{Berry}} > 0$ and $D_{\text{Metric}} > 0$, which is visualized in a tilted gapped Dirac band structure in Fig. 3, E and G. We now flip the AFM order to the AFM II by performing

time reversal T. Under T, the bands are flipped between $\pm\mathbf{k}$ (Fig. 3, F and H), the Berry curvature flips sign, but the quantum metric keeps the same sign. Hence, from Fig. 3, F and H, one can see that, $D_{\mathrm{Berry}}(\mathrm{AFMII}) = D_{\mathrm{Berry}}(\mathrm{AFMI})$, but $D_{\mathrm{Metric}}(\mathrm{AFMII}) = -D_{\mathrm{Metric}}(\mathrm{AFMI})$. Therefore, our observation that the nonlinear Hall signal flips sign upon reversing the AFM order excludes the Berry curvature dipole mechanism. In section II.1 of (*19*), we enumerate experimental results, including the relation with AFM order, scaling, vertical electric field dependence, and relation with mirror symmetry, which corroborate that the Berry curvature dipole mechanism cannot account for our data.

Within the nonlinear effects that flip sign upon reversing the AFM order, there is another possibility, the second-order Drude effect (*5*, *12*, *17*, *22*). This effect is expected to be proportional to τ^2 (*22*) and therefore can be ruled out on the basis of our scaling data in Fig. 3A. Moreover, the NHE is antisymmetric (upon exchanging the first two indices), $\sigma^{\mathrm{NHE}}_{\alpha\beta\gamma} = -\sigma^{\mathrm{NHE}}_{\beta\alpha\gamma}$, but the second-order Drude effect (SODE) is symmetric, $\sigma^{\mathrm{SODE}}_{\alpha\beta\gamma} = \sigma^{\mathrm{SODE}}_{\beta\alpha\gamma}$ (*22*). Using an electrical sum-frequency generation method [section II.2 of (*19*)], we showed that our signal is indeed antisymmetric, $\sigma^{2\omega}_{yxx} = -\sigma^{2\omega}_{xyx}$, which demonstrates that the SODE is insignificant in our signal. In section II.2.3 of (*19*), we present additional data that suggest that the NHE is dominant over the SODE at different temperatures and charge densities. In section III.5 of (*19*), we show that the nonlinear Hall signal is negligibly small at ±8 T because the forced ferromagnetic state recovers inversion symmetry. We also carefully addressed other competing origins such as thermal and accidental diode junctions [section II.3 of (*19*)].

Energy-resolved probe of quantum metric in *PT*-symmetric antiferromagnets

We also study the evolution of the nonlinear conductivity $\sigma^{2\omega}_{yxx}$ with the charge density. The nonlinear Hall signal is zero inside the charge neutrality gap (Fig. 4A). This is consistent with the expectation that the NHE is a Fermi surface property. As we tune the Fermi energy away from the charge neutrality, the nonlinear Hall signal emerges. Notably, the conductivities in electron and hole regimes have the same sign. As we go deeper into the electron-doped regime, the signal reverses sign.

We now provide an intuitive physical picture to understand the large quantum metric dipole and its Fermi-level dependence. $MnBi_2Te_4$ features Dirac surface states, which are gapped owing to the AFM, leading to a large quantum metric near the gap edge. Moreover, because the AFM order breaks both T and P, the Dirac bands are asymmetric about $\mathbf{k} = 0$ (Fig. 3G). Hence, at a fixed energy, positive and negative momenta have a different quantum metric, leading to a nonzero quantum metric dipole.

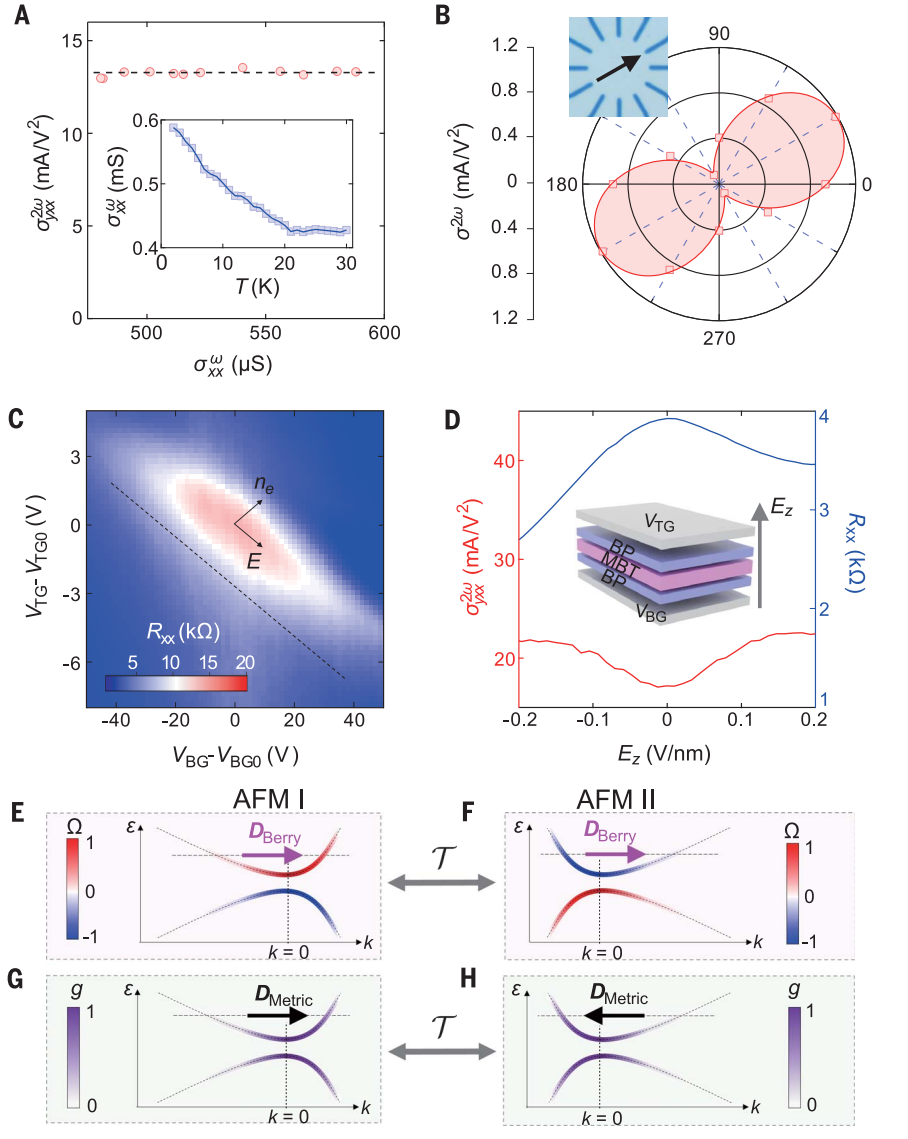


Fig. 3. Systematic investigation of the NHE. (**A**) The scaling between the nonlinear Hall conductivity and the Drude conductivity $\sigma_{xx} = 1/\mathrm{R}_{xx}$. The nonlinear Hall conductivity can be directly extracted from the data as $\sigma^{2\omega}_{yxx} = J^{2\omega}_{yxx}/E^{\omega 2}_{x} = \left(V^{2\omega}_{yxx}/3\right)/\left(I^{\omega 2}_{x}R^{3}_{xx}w^{2}d\right)$. (**B**) Angular dependence of the nonlinear Hall conductivity in Device-BM1. (**C**) Dual-gated resistance map of the 2L BP/6SL $MnBi_2Te_4$/2L BP heterostructure (Device-BMB1). The vertical electric field E_z and charge density dependence can be independently tuned by combining the top and bottom gate voltages. (**D**) E_z dependence of the nonlinear Hall conductivity and linear longitudinal resistance. E_z follows the dashed line in (C). (**E** to **H**) Schematic illustration of the Berry curvature dipole (D_{Berry}) and the quantum metric dipole (D_{Metric}) for the AFM I and AFM II of the BP/6SL $MnBi_2Te_4$/BP heterostructure (see text). Although we aligned the crystalline axes of BP and $MnBi_2Te_4$ in our Device-BMB1 (Fig. 2A), realistically it is difficult to make the alignment perfect. If the alignment is imperfect and *PT* symmetry is broken, a Berry curvature dipole is allowed.

Intuitively, we can understand the sign of the nonlinear Hall signal by considering whether positive or negative momenta have a larger quantum metric. We see from Fig. 3G that both upper and lower parts of the Dirac cone have $g(+k_{\mathrm{F}}) > g(-k_{\mathrm{F}})$, suggesting that the nonlinear Hall signals should show the same sign in electron and hole regimes, consistent with our data (Fig. 4A). The sign change in the electron-doped regime is beyond this simple picture.

To achieve a more comprehensive understanding, we built an effective model of the BP/6SL $MnBi_2Te_4$/BP heterostructure [section IV.4-9 of (*19*)]. Owing to the incommensurability of the BP and $MnBi_2Te_4$ lattices, we need to derive the coupling between the Bloch states of the two materials in the real-space continuum [i.e., within the extended Brillouin zone (BZ)]. The low-energy bands are located in the BZ center, so only Bloch

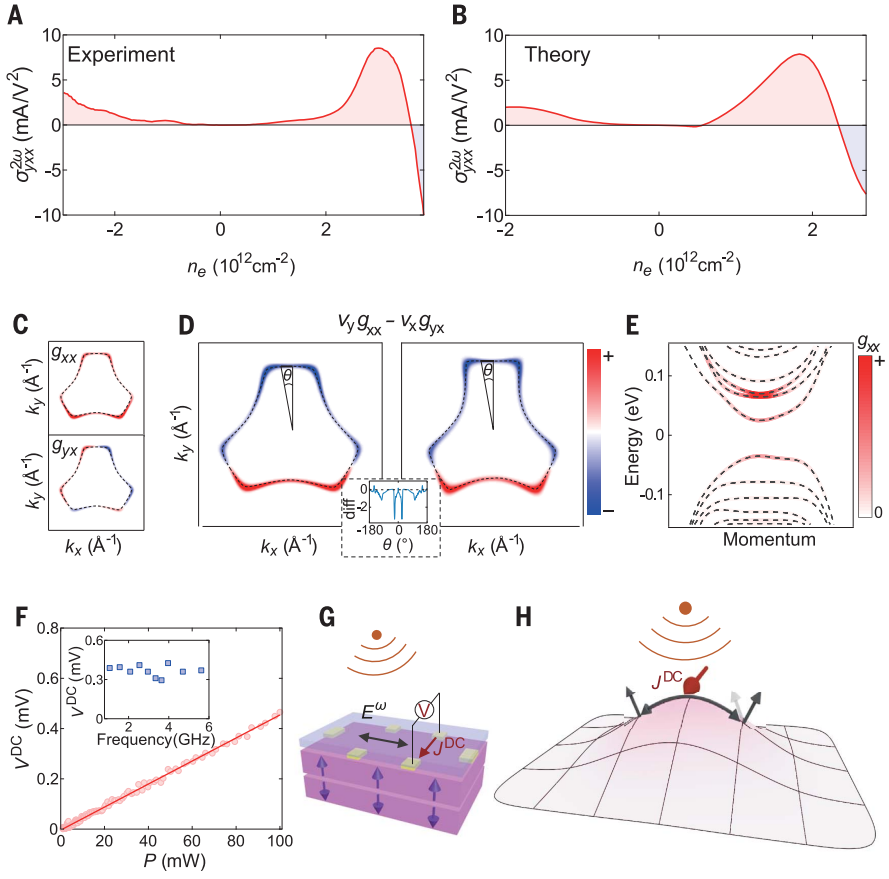


Fig. 4. The quantum metric dipole as the microscopic geometrical origin. (**A**) Experimentally measured nonlinear Hall conductivity as a function of charge density. (**B**) Theoretically calculated $\sigma^{2\omega}_{yxx}$ as a function of n_e based on the BP/6SL $MnBi_2Te_4$/BP band structure (see text). (**C** to **E**) The electronic structure of the BP/6SL $MnBi_2Te_4$/BP heterostructure calculated with an effective model (see text). (C) Fermi surface at −50 meV (the lower part of the surface Dirac cone). The coupling between $MnBi_2Te_4$ and BP is turned off, so that contour respects C_{3z} symmetry. The quantum metric g_{xx} and g_{yx} plotted around the Fermi surface. (D) The nonlinear Hall conductivity is given by the integral $D_{\text{Metric}} = \int_k (v_y g_{xx} - v_x g_{yx})\delta(\varepsilon - \varepsilon_F)$ around the Fermi surface. With C_{3z} symmetry (left panel), the integral goes to zero. After turning on the coupling between $MnBi_2Te_4$ and BP (right panel), C_{3z} is broken, making the integral around the Fermi contour nonzero. To more clearly see how the integral changes to nonzero when C_{3z} is broken, we rewrite D_{Metric} as an integral of the polar angle θ, $D_{\text{Metric}} = \int_{FS} (v_y g_{xx} - v_x g_{yx}) \frac{dl}{d\theta} d\theta$ (dl is an infinitesimal length along the Fermi surface). The inset presents the change of the above kernel [see section III.16 of (*19*) for details]. (E) Band structure of BP/6SL $MnBi_2Te_4$/BP heterostructure as a function of k_y for k_x = 0. Color represents the quantum metric of the bands. (**F**) Measured microwave rectification based on the intrinsic NHE. Inset: DC signal as a function of microwave frequency. (**G**) Schematic illustration of microwave rectification. (**H**) Schematic illustration of quantum metric–induced nonlinear responses. The horizontal axes are k_x and k_y. The two black arrows represent the Bloch wave functions at two nearby k points. The two arrows point to different directions, illustrating a finite distance between two wave functions (i.e., a finite quantum metric). This quantum metric leads to a NHE, which can turn an external AC electric field (e.g., the microwave in the figure) into a DC signal.

bands with the same momentum hybridize. The coupling amplitude depends only on the characteristic decay length of the atomic orbitals as any discrete lattice structure is averaged out (*47*). The Hamiltonian reads $\hat{h}(k_x, k_y) = \begin{pmatrix} \hat{h}_{\text{MBT}} & \hat{U}_t & \hat{U}_b \\ \hat{U}_t^{\dagger} & \hat{h}_{\text{BP},\,t} & 0 \\ \hat{U}_b^{\dagger} & 0 & \hat{h}_{\text{BP},\,b} \end{pmatrix}$. $\hat{h}_{\text{MBT}}$ and $\hat{h}_{\text{BP},\,t(b)}$ are Hamiltonians for 6SL $MnBi_2Te_4$ and top (bottom) BP. The spin-orbit coupling (SOC) in $MnBi_2Te_4$ is crucial for a nonzero NHE because it allows the low-energy orbitals to feel the symmetry breaking by the AFM order [section III.15 of (*19*)]. In particular, the SOC was included in the model following the original work by (*50*). $\hat{U}_t$ and $\hat{U}_b$ denote the band hybridization caused by nearest-neighbor coupling between $MnBi_2Te_4$ and BP. The bare $MnBi_2Te_4$ and BP parameters were obtained by fitting the first-principle band structures. The $MnBi_2Te_4$ and BP coupling parameters were partly constrained by considering the independent data of R_{xx}/R_{yy} so that an agreement in the overall magnitude was achieved independently. We adjusted the remaining free coupling model parameters to match detailed features in the charge density dependence [section IV.9 of (*19*)].

We first turn off the coupling between the $MnBi_2Te_4$ and BP. The Fermi surface shown in Fig. 4C (−50 meV) is C_{3z} symmetric, and there is already a large quantum metric (g_{xx} and g_{yx}) around it. According to (*22*), the D_{Metric} responsible for the nonlinear Hall is given by $D_{\text{Metric}} = \int_{\mathbf{k}} (v_y g_{xx} - v_x g_{yx})\delta(\varepsilon - \varepsilon_F)$ (v is the Fermi velocity). We plot the integral kernel in color in Fig. 4D. Positive and negative contributions around the contour exactly cancel each other out because of C_{3z} symmetry. So, the integral goes to zero (Fig. 4D, left panel). We then turn on the $MnBi_2Te_4$-BP couplings, which breaks C_{3z}. For the C_{3z}-breaking contour, we observe unequal contributions from the two colors, leading to a nonzero D_{Metric} [Fig. 4D, right panel; see details in the caption and in section III.16 of (*19*)]. Figure 4E shows the band structure of the BP/6SL $MnBi_2Te_4$/BP heterostructure, from which we can compute the intrinsic nonlinear Hall conductivity $\sigma^{2\omega}_{yxx}$ as a function of chemical potential. In particular, near the charge neutrality gap, we found that $\sigma^{2\omega}_{yxx}$ indeed mainly comes from the quantum metric of the Dirac surface states, consistent with the intuitive picture above. The sign inversion in the electron-doped regime mainly comes from the quantum metric of the avoided crossing inside conduction bands according to our calculation [section IV.9 of (*19*)]. Owing to the multiband nature of our model, the $\sigma^{2\omega}_{yxx}$ was calculated by the general expression $\sigma^{2\omega}_{yxx} = -2e^3 \sum_{n,m}^{\varepsilon_m \neq \varepsilon_n} \text{Re} \int_{\mathbf{k}} \left(\frac{v_y^n \langle u_n | i\partial_{k_x} u_m \rangle \langle u_m | i\partial_{k_x} u_n \rangle}{\varepsilon_n - \varepsilon_m} - \frac{v_x^n \langle u_n | i\partial_{k_y} u_m \rangle \langle u_m | i\partial_{k_x} u_n \rangle}{\varepsilon_n - \varepsilon_m} \right) \delta(\varepsilon_n - \varepsilon_F)$ (*22*). This general expression can be decomposed into the quantum metric dipole contribution plus additional interband contributions (AIC)

$$\sigma^{2\omega}_{yxx} = -2e^3 \sum_n \int_{\mathbf{k}} \frac{v_y^n g_{xx}^n - v_x^n g_{yx}^n}{\varepsilon_n - \varepsilon_{\bar{n}}} \delta(\varepsilon_n - \varepsilon_F) + \text{AIC} \quad (1)$$

where the first term is the quantum metric dipole contribution, and the second term is $\text{AIC} = -2e^3 \sum_{n,m}^{\varepsilon_m \neq \varepsilon_n, \varepsilon_{\bar{n}}} \text{Re} \int_{\mathbf{k}} \left(\frac{v_y^n \langle u_n | i\partial_{k_x} u_m \rangle \langle u_m | i\partial_{k_x} u_n \rangle}{\varepsilon_n - \varepsilon_m} - \frac{v_x^n \langle u_n | i\partial_{k_y} u_m \rangle \langle u_m | i\partial_{k_x} u_n \rangle}{\varepsilon_n - \varepsilon_m} \right) \frac{\varepsilon_m - \varepsilon_{\bar{n}}}{\varepsilon_n - \varepsilon_m} \delta(\varepsilon_n - \varepsilon_F)$ ($\bar{n}$ is the band whose energy is closest to n). In our BP/6SL $MnBi_2Te_4$/BP system, we found that the quantum metric dipole contribution strongly dominates, whereas the AIC is small [details

in section IV.3 of (*19*)]. Therefore, our nonlinear Hall measurement is a powerful, energy-resolved probe of the quantum metric.

Possible AFM spin-based wireless rectification

The second-order nonlinear effect enables not only frequency doubling but also rectification. We use the intrinsic AFM nonlinear Hall effect to demonstrate wireless rectification with zero external bias (battery-free) and without magnetic field. We inject microwave radiation and measure the DC signal. We observe clear rectification DC voltage in response to the microwave radiation (Fig. 4F), which shows a broad band response, including the Wi-Fi frequencies (2.4 and 5 GHz) and even higher frequencies [see fig. S35 and section V.2 of (*19*)]. In section III.12 of (*19*), we show that the rectification signal flips sign as we reverse the AFM state, which suggests the intrinsic quantum metric dipole origin. Apart from the intrinsic quantum metric dipole, extrinsic sources such as the Schottky diodes at the metal-$MnBi_2Te_4$ junction, unintentional diodes inside the $MnBi_2Te_4$, and the gap between the two gates can also lead to microwave rectification. To unambiguously rule out these extrinsic sources, future systematic experiments will be needed [discussion in section V.1 of (*19*)].

Discussion and outlook

The intrinsic second-order Hall effect observed here realizes an electrical nonlinearity induced by the AFM spins and provides a rare example of a quantum metric response. As highlighted by recent theoretical studies, the influence of the quantum metric is expected to span many different areas, ranging from nonlinear responses in *PT*-symmetric AFMs to flat-band conductivity, superconductivity and charge orders in moiré systems, the fractional Chern insulator, and **k**-space dual of gravity (*20*, *22*–*34*). Another interesting future direction is to explore the nonlinear responses in canted AFM materials [section V.3 of (*19*)], where nonzero Berry curvature of higher order in magnetization has recently been observed (*38*, *46*, *51*). In terms of materials, we show that, beyond "band structure engineering," the van der Waals interfaces can engineer the properties of the wave function i.e., "quantum geometry engineering" (*47*). Our observations may enable the use of AFM spins to harvest electromagnetic energy and to realize self-powered AFM spintronic devices. An exciting future breakthrough would be to demonstrate room-temperature wireless rectification based on the quantum metric NHE in a *PT*-symmetric AFM material.

REFERENCES AND NOTES

1. J. Provost, G. Vallee, *Commun. Math. Phys.* **76**, 289–301 (1980).
2. Y. Tokura, N. Nagaosa, *Nat. Commun.* **9**, 3740 (2018).
3. J. E. Moore, J. Orenstein, *Phys. Rev. Lett.* **105**, 026805 (2010).
4. I. Sodemann, L. Fu, *Phys. Rev. Lett.* **115**, 216806 (2015).
5. T. Ideue *et al*., *Nat. Phys.* **13**, 578–583 (2017).
6. K. Kang, T. Li, E. Sohn, J. Shan, K. F. Mak, *Nat. Mater.* **18**, 324–328 (2019).
7. Q. Ma *et al*., *Nature* **565**, 337–342 (2019).
8. S. Dzsaber *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2013386118 (2021).
9. D. Kumar *et al*., *Nat. Nanotechnol.* **16**, 421–425 (2021).
10. K. Yasuda *et al*., *Nat. Nanotechnol.* **15**, 831–835 (2020).
11. W. Zhao *et al*., *Nat. Mater.* **19**, 503–507 (2020).
12. H. Isobe, S.-Y. Xu, L. Fu, *Sci. Adv.* **6**, eaay2497 (2020).
13. S. Lai *et al*., *Nat. Nanotechnol.* **16**, 869–873 (2021).
14. P. He *et al*., *Nat. Nanotechnol.* **17**, 378–383 (2022).
15. Z. Zhang *et al*., *Nat. Commun.* **13**, 6191 (2022).
16. S. Sinha *et al*., *Nat. Phys.* **18**, 765–770 (2022).
17. N. J. Zhang *et al*., Momentum polarized superconductivity in twisted trilayer graphene. Preprint at https://arxiv.org/abs/2209.12964 (2022).
18. Z. Sun *et al*., *Nature* **572**, 497–501 (2019).
19. See additional text and data in the supplementary materials.
20. Y. Gao, S. A. Yang, Q. Niu, *Phys. Rev. Lett.* **112**, 166601 (2014).
21. A. Gianfrate *et al*., *Nature* **578**, 381–385 (2020).
22. C. Wang, Y. Gao, D. Xiao, *Phys. Rev. Lett.* **127**, 277201 (2021).
23. H. Liu *et al*., *Phys. Rev. Lett.* **127**, 277202 (2021).
24. S. Lahiri, K. Das, D. Culcer, A. Agarwal, Intrinsic nonlinear conductivity induced by the quantum metric dipole. Preprint at https://arxiv.org/abs/2207.02178 (2022).
25. T. B. Smith, L. Pullasseri, A. Srivastava, *Phys. Rev. Res.* **4**, 013217 (2022).
26. A. Arora, M. S. Rudner, J. C. W. Song, *Nano Lett.* **22**, 9351–9357 (2022).
27. J. Mitscherling, T. Holder, *Phys. Rev. B* **105**, 085154 (2022).
28. J.-W. Rhim, K. Kim, B.-J. Yang, *Nature* **584**, 59–63 (2020).
29. P. J. Ledwith, G. Tarnopolsky, E. Khalaf, A. Vishwanath, *Phys. Rev. Res.* **2**, 023237 (2020).
30. T. Holder, D. Kaplan, B. Yan, *Phys. Rev. Res.* **2**, 033100 (2020).
31. H. Watanabe, Y. Yanase, *Phys. Rev. X* **11**, 011001 (2021).
32. K.-E. Huhtinen, J. Herzog-Arbeitman, A. Chew, B. A. Bernevig, P. Törmä, *Phys. Rev. B* **106**, 014518 (2022).
33. J. S. Hofmann, E. Berg, D. Chowdhury, *Phys. Rev. Lett.* **130**, 226001 (2023).
34. X. Hu, T. Hyart, D. I. Pikulin, E. Rossi, *Phys. Rev. B* **105**, L140506 (2022).
35. N. Nagaosa, J. Sinova, S. Onoda, A. MacDonald, N. P. Ong, *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
36. M. M. Otrokov *et al*., *Nature* **576**, 416–422 (2019).
37. D. Zhang *et al*., *Phys. Rev. Lett.* **122**, 206401 (2019).
38. S. H. Lee *et al*., *Phys. Rev. Res.* **1**, 012011 (2019).
39. Y. Deng *et al*., *Science* **367**, 895–900 (2020).
40. C. Liu *et al*., *Nat. Mater.* **19**, 522–527 (2020).
41. H. Deng *et al*., *Nat. Phys.* **17**, 36–42 (2021).
42. D. Ovchinnikov *et al*., *Nano Lett.* **21**, 2544–2550 (2021).
43. S. Yang *et al*., *Phys. Rev. X* **11**, 011003 (2021).
44. A. Gao *et al*., *Nature* **595**, 521–525 (2021).
45. L. Tai *et al*., *ACS Nano* **16**, 17336–17346 (2022).
46. S.-K. Bac *et al*., *npj Quantum Mater.* **7**, 46 (2022).
47. T. Akamatsu *et al*., *Science* **372**, 68–72 (2021).
48. A. Iyama, T. Kimura, *Phys. Rev. B* **87**, 180408 (2013).
49. S. Jiang, J. Shan, K. F. Mak, *Nat. Mater.* **17**, 406–410 (2018).
50. H. Zhang *et al*., *Nat. Phys.* **5**, 438–442 (2009).
51. J. Kipp *et al*., *Commun. Phys.* **4**, 99 (2021).
52. A. Gao, S. Xu, Quantum metric nonlinear Hall effect in a topological antiferromagnetic heterostructure, Zenodo (2023); https://doi.org/10.5281/zenodo.8002209.

ACKNOWLEDGMENTS

We are grateful to A. Yacoby for help with the microwave measurements. We thank M. Wesson and N. Poniatowski for technical support during the microwave measurements. We also thank Y. Gao, J. Ahn, and P. Kim for helpful discussions. We are grateful to L. Ye, M. Mogi, Y. Fujishiro, and T. Kurumaji for extensive discussions on the scaling of AHE. **Funding:** Work in the S.-Y.X. group was partly supported through the Center for the Advancement of Topological Semimetals (CATS), an Energy Frontier Research Center (EFRC) funded by the US Department of Energy (DOE) Office of Science (fabrication and measurements), through the Ames National Laboratory under contract DE-AC0207CH11358, and partly through Air Force Office of Scientific Research (AFOSR) grant FA9550-23-1-0040 (data analysis and manuscript writing). S.-Y.X. acknowledges the Corning Fund for Faculty Development. Q.M. and L.F. acknowledge support from the NSF Convergence program (NSF ITE-2235945) and the CIFAR program. S.-Y.X. and D.B. were supported by the NSF Career DMR-2143177. C.T. and Z.S. acknowledge support from the Swiss National Science Foundation under project P2EZP2 191801 and P500PT 206914, respectively. Y.F.L., S.-Y.X., D.C.B., Y.O., and L.F. were supported by the STC Center for Integrated Quantum Materials (CIQM), NSF grant no. DMR-1231319. This work was performed in part at the Center for Nanoscale Systems (CNS) Harvard University, a member of the National Nanotechnology Coordinated Infrastructure Network (NNCI), which is supported by the National Science Foundation under NSF award no. 1541959. Bulk single-crystal growth and characterization of $MnBi_2Te_4$ were performed at UCLA and were supported by the DOE, Office of Science, under award no. DE-SC0021117. The work at Northeastern University was supported by the Air Force Office of Scientific Research under award no. FA9550-20-1-0322, and it benefited from the computational resources of Northeastern University's Advanced Scientific Computation Center (ASCC) and the Discovery Cluster. The work in the QM group was partly supported through the CATS, an EFRC funded by the DOE Office of Science, through the Ames National Laboratory under contract DE-AC0207CH11358 (fabrication and measurements) and partly through NSF DMR-2143426 (data analysis and manuscript writing). T.V.T. and P.P.O. were supported by the CATS, an EFRC funded by the DOE Office of Science, through the Ames National Laboratory under contract DE-AC0207CH11358. T.R.C. was supported by the 2030 Cross-Generation Young Scholars Program of the National Science and Technology Council (NSTC) in Taiwan (program no. MOST111-2628-M-006-003-MY3); National Cheng Kung University (NCKU), Taiwan; and the National Center for Theoretical Sciences (NCTS), Taiwan. This research was supported, in part, by the Higher Education Sprout Project, Ministry of Education to the Headquarters of University Advancement at NCKU. H.L. acknowledges support from the National Science and Technology Council (NSTC) in Taiwan under grant no. MOST 111-2112-M-001-057-MY3. The work at TIFR Mumbai was supported by the Department of Atomic Energy of the Government of India under project no. 12-R&D-TFR-5.10-0100 and benefited from the computational resources of TIFR Mumbai. K.W. and T.T. acknowledge support from the JSPS KAKENHI (grant nos. 20H00354, 21H05233, and 23H02052) and World Premier International Research Center Initiative (WPI), MEXT, Japan. M.H. and C.R.D. were supported by the AFOSR under award no. FA9550-20-1-0319. S.W.C. acknowledges partial support from the Harvard Quantum Initiative in Science and Engineering. **Author contributions:** S.-Y.X. conceived the experiments and supervised the project. A.G. fabricated the devices, performed the measurements, and analyzed data with help from Y.F.L., D.B., J.X.Q., H.C.L., C.T., T.D., Z.S., S.C.H., D.C.B., and Q.M. A.G. and S.W.C. performed the microwave rectification experiments. C.H., T.Q., and N.N. grew the $MnBi_2Te_4$ single crystals. M.H. and C.R.D. performed NV center magnetometry. B.G. made the theoretical studies, including first-principles calculations and effective modeling, with the help from T.V.T., Y.O., H.J.T., S.W.L., B.S., H.L., A.B., T.R.C., L.F., and P.P.O. T.V.T. developed the effective model with help from B.G. under the guidance of P.P.O. K.W. and T.T. grew the hBN single crystals. S.-Y.X., A.G., and Q.M. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Data in the main text and supplementary materials, as well as the codes for theoretical calculations of the quantum metric, are available from Zenodo (*52*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.sciencemag.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf1506
Materials and Methods
Supplementary Text
Figs. S1 to S50
Tables S1 to S3
References (*53*–*71*)

Submitted 5 October 2022; accepted 6 June 2023
Published online 15 June 2023
10.1126/science.adf1506

CHATGPT

Experimental evidence on the productivity effects of generative artificial intelligence

Shakked Noy* and Whitney Zhang

We examined the productivity effects of a generative artificial intelligence (AI) technology, the assistive chatbot ChatGPT, in the context of midlevel professional writing tasks. In a preregistered online experiment, we assigned occupation-specific, incentivized writing tasks to 453 college-educated professionals and randomly exposed half of them to ChatGPT. Our results show that ChatGPT substantially raised productivity: The average time taken decreased by 40% and output quality rose by 18%. Inequality between workers decreased, and concern and excitement about AI temporarily rose. Workers exposed to ChatGPT during the experiment were 2 times as likely to report using it in their real job 2 weeks after the experiment and 1.6 times as likely 2 months after the experiment.

Recent advances in generative artificial intelligence (AI) may have widespread implications for production and labor markets. Generative AI systems such as ChatGPT or DALL-E, which can be prompted to create new text or visual outputs from large amounts of training data, are qualitatively unlike most historical examples of automation technologies. Previous waves of automation predominantly affected "routine" tasks consisting of explicit sequences of steps that could be easily codified and programmed into a machine or computer, such as assembly-line manufacturing tasks or bookkeeping tasks (*1*, *2*). By contrast, creative, difficult-to-codify tasks such as writing and image generation avoided automation, a pattern scholars have noted might change with the advent of deep learning, which now underpins generative AI systems.

The emergence of powerful generative AI technologies reintroduces a host of classic questions in a new context (*3*–*5*). Automation technologies by definition perform specific tasks in place of humans. But more broadly, these technologies may either displace humans completely from certain occupations or augment existing human workers by increasing their productivity (*6*–*9*). If automation technologies such as industrial robots mostly displace human workers, they can increase unemployment. Moreover, their impacts on aggregate productivity may be small or nonexistent to the degree that they mainly serve to redistribute income previously earned by displaced workers to the capital owners supplying their robot replacements (*10*). If automation technologies such as computers augment existing workers, they can simultaneously benefit workers, capital owners, and consumers by raising wages, boosting productivity, and lowering prices (*11*–*13*).

A potent generative writing tool such as ChatGPT could conceivably either displace or augment human labor. ChatGPT could entirely replace certain kinds of writers, such as grant writers or marketers, by letting companies directly automate the creation of grant applications and press releases with minimal human oversight. Alternatively, instead of displacing workers, ChatGPT could substantially raise the productivity of grant writers and marketers, for example, by automating relatively routine, time-consuming subcomponents of their writing tasks, such as translating ideas into a rough draft. In this case, these services would become cheaper and demand could expand, resulting in higher employment and greater productivity for companies, cheaper products for consumers, and potentially higher wages for workers (*14*). Furthermore, inequalities between workers could either decrease if lower-ability workers are supported more by ChatGPT or increase if higher-ability workers have the skills necessary to take advantage of the new technology.

Which of these eventualities will generative AI systems bring about? The answer depends on a host of research questions (RQs). RQ1: How does access to generative AI systems affect workers' productivity in existing tasks? Do workers choose to use these systems? Conditional on using these systems, how do workers interact with them and how do they affect productivity (*15*–*18*)? RQ2: Do these systems differentially affect low- and high-ability workers? RQ3: How do workers subjectively react to these technologies (*19*)?

Methods

This paper took the first step toward answering these questions (*20*). In a preregistered online experiment, we recruited 453 experienced, college-educated professionals on the survey platform Prolific and assigned each to complete two occupation-specific, incentivized writing tasks (*21*). The experiment took place from 27 January to 21 February 2023 and involved GPT-3.5. The occupations that we drew from were marketers, grant writers, consultants, data analysts, human resource professionals, and managers. The tasks, which included writing press releases, short reports, analysis plans, and delicate emails, comprised 20- to 30-min assignments designed to resemble real tasks performed in these occupations. Indeed, most of our participants reported completing similar tasks before and rated the assigned tasks as realistic representations of their everyday work (see the supplementary materials).

Participants faced high-powered incentives in the form of large bonus payments to produce high-quality work. They received a base payment of $10 plus up to $14 in bonus payments for output quality, with the average overall rate of $17/hour substantially exceeding the Prolific standard of $12/hour. We cross-randomized the structure of bonus payments faced by participants to show the robustness of our results to different incentive schemes (see below for more details). Output quality was assessed by blinded experienced professionals working in the same occupations. Evaluators were asked to treat the output as if encountered in a work setting and were incentivized to grade outputs carefully on a scale of 1 to 7 (*22*). Each piece of output was seen by three evaluators, with an average within-essay cross-evaluator correlation of 0.44 (*23*).

We randomly assigned 50% of participants to the treatment group and 50% to the control group. The treatment group was instructed to register for ChatGPT between the first and second task, received guidance on using it, and were told they were permitted to use it on the second task if they found it useful. The control group was instead instructed to register for the LaTeX editor Overleaf in an attempt to hold the time and hassle costs of signup constant between the two groups. The control group was not told they could use Overleaf on the second task and <5% of participants subsequently reported using it.

In addition to output quality evaluations, we collected self-reported and objective measures of participants' time spent on the tasks and took snapshots of participants' outputs each minute while they performed the task to construct objective measures of activity and to detect ChatGPT usage (see the supplementary materials).

A complete description of our experimental design, copies of relevant survey questionnaires, and additional figures validating our central measures and extending our main results are included in the supplementary materials. Descriptive statistics about the sample, as well as balance and selective attrition tests, are

Department of Economics, Massachusetts Institute of Technology, Cambridge, MA, USA.
*Corresponding author. Email: snoy@mit.edu

available in Table 1. The attrition rate was 6% in the control group and 11% in the treatment group. Balance tests indicate that across 13 pretreatment characteristics, the treatment and control groups exhibited a small but significant difference only for only two characteristics: employment status and being a human resources professional. Our partly within-person design, which controls for performance on the pretreatment task, should eliminate any influence of selective attrition on our results. In the supplementary materials, we also report Lee bounds (*24*) on our main results and versions of our results controlling for employment status and occupation, which confirmed that our results are highly robust to selective attrition.

Results

Take-up of ChatGPT

In the treatment group, 92% of treated participants successfully registered for ChatGPT, and 80% chose to use it on the second task (*25*). Users gave it an average self-assessed usefulness score of 4.4 out of 5.

Before treatment, 70% of our participants had heard of ChatGPT and 32% had used it before. Self-reported and objective measures indicated that ~10 to 20% of the control group used ChatGPT on the tasks (see the supplementary materials), meaning there was at least a 60-percentage point experimentally induced gap in usage between our treatment and control groups on the second task. Our estimates reflect the effects of ChatGPT on the average productivity of the 60 to 70% of participants whose usage was determined by their treatment assignment, and constitute lower bounds on the effects of ChatGPT usage on productivity. In the supplementary materials, we report two-stage least-squares results adjusting our estimates upward for imperfect compliance.

Productivity

We first show results for our two productivity measures: time taken and evaluator grades (Fig. 1). The experimental intervention shifted both outcomes substantially. In the treatment group, time taken on the posttreatment task dropped by 11 min (0.75 SDs) relative to the control group, who took an average of 27 min ($P < 0.001$). Average evaluator grades in the treatment group increased by 0.45 SDs ($P < 0.001$), with similar increases for overall grades and specific grades for writing quality, content quality, and originality.

These effects are not limited to specific pockets of the time or grade distributions. As shown in Fig. 1, C and D, the entire time distribution shifted to the left (faster work) and the entire grade distribution shifted to the right (higher quality). At the individual worker level, as shown in Fig. 2, treated workers who received a low grade on the first task experienced both 1- to 2-point increases in grades and 10-min decreases in time spent, whereas workers who received a high grade maintained their grade level while also reducing their time spent by ~10 min.

These results are virtually identical across our two main incentive schemes, which covered 80% of respondents: a "linear" scheme in which respondents were paid $1 for each point they received on each submission (each of which was graded on a 1- to 7-point scale), and a "convex" scheme in which respondents were additionally paid $3 for earning a grade of 6 or 7. The results shown in Fig. 1 are based on these two incentive schemes. The fact that treated participants reduced their time spent by a similar amount even when faced with strong incentives to produce high-quality output (under the convex scheme) demonstrates that the time-saving effects of ChatGPT are not specific to linear payment regimes and apply robustly across incentive structures.

In our third incentive arm involving 20% of participants, we required participants to spend exactly 15 min on each task, thereby holding effort fixed across the treatment and control groups and allowing us to interpret any difference in grades as a pure effect of ChatGPT access on productive capacity. In this arm, the treatment increased grades by a similar albeit not statistically significant 0.33 SDs (*26*).

In an additional intervention, after completing the second task, 30% of the treatment group were shown their first-task human-created output and given the opportunity to edit or replace it using ChatGPT. Of these participants, 19% chose to replace their response with ChatGPT's output and another 17% used ChatGPT to edit their original response, suggesting that participants viewed ChatGPT as a means to improve output quality as well as save time.

Productivity inequality

The control group exhibited persistent productivity inequality: Participants who scored well on the first task also tended to score well on the second task. As Fig. 2A shows, there was a correlation of 0.41 ($P < 0.001$) between a control participant's grade on the first task and their grade on the second task, holding the evaluator constant.

In the treatment group, initial inequalities were more than half-erased by the treatment: The correlation between first-task and second-task grades was only 0.14 ($P < 0.001$ for the difference in slopes). This reduction

Table 1. Descriptive statistics.

Variable	*N* (control)	Mean (control)	*N* (treatment)	Mean (treatment)	Difference
Annual salary in main job ($)	234	67,764	213	71,938	4173
Years of tenure in occupation	234	10.49	215	10.07	−0.43
Employed	226	91%	210	96%	5.0%**
Occupation: HR professional	235	6%	218	11%	4.6%*
Occupation: business consultant	235	13%	218	11%	−1.3%
Occupation: data analyst	235	11%	218	11%	−0.0%
Occupation: grant writer	235	16%	218	17%	1.2%
Occupation: manager	235	43%	218	41%	−1.7%
Occupation: marketer	235	11%	218	9%	−2.3%
Time spent (task 1, min)	227	26.10	212	26.58	0.47
Average grade (task 1)	233	3.63	211	3.77	0.15
Job satisfaction (task 1, scale of 1 to 10)	234	6.30	215	6.34	0.04
Self-efficacy (task 1, scale of 1 to 10)	234	6.89	215	6.90	0.01

This table presents descriptive statistics for our sample. We recode salary reports of >$500,000 to missing (affects two observations). "Employed" includes full-time and part-time employment. *$P < 0.10$; **$P < 0.05$.

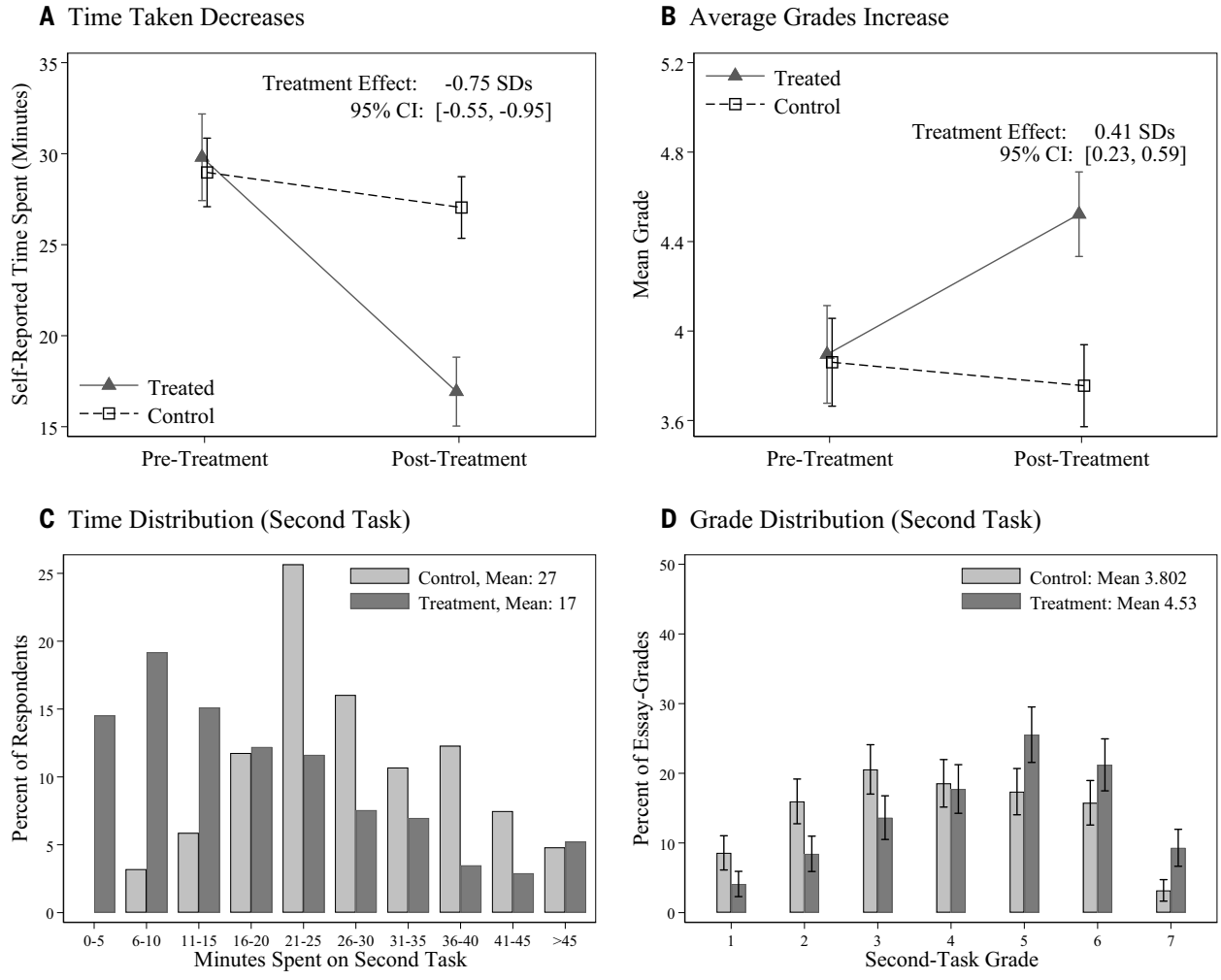


Fig. 1. Treatment effects on productivity. Effects shown are restricted to the linear and convex incentive groups. (**A**) and (**B**) Means (and 95% confidence intervals for those means) of self-reported time taken and average grades in the first and second task separately in the treatment and control groups. The results look very similar for the objective measure of time active (see the supplementary materials). Also shown are treatment effect coefficients and 95% confidence intervals, rescaled to be in terms of pretreatment SDs of the outcome variable. The coefficients are estimated from regressions of the within-participant change in outcome from before to after treatment on a treatment dummy, occupation*task order fixed effects, and incentive arm fixed effects. In (A), this is at the participant level and SEs are heteroskedasticity robust. In (B), this is at the participant-evaluator level, the regression also includes grader fixed effects, and SEs are clustered at the participant level. (**C** and **D**) Raw graphs of the outcome distribution in the treatment versus control group on the second task; (C) is at the participant level and (D) is at the participant-evaluator level.

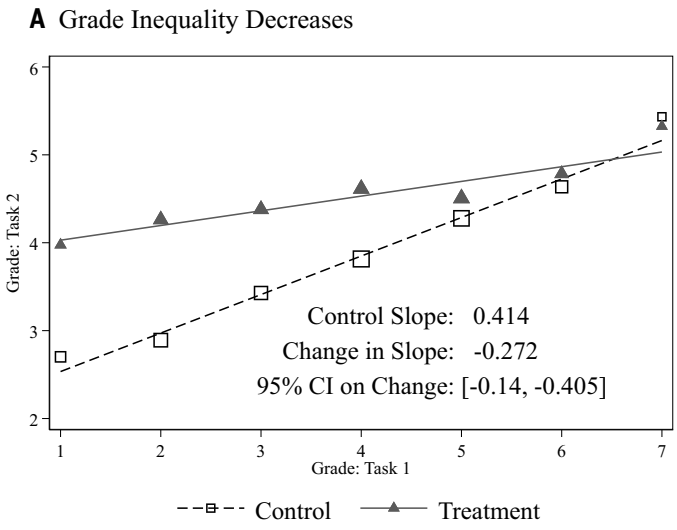
in inequality was driven by the fact that participants who scored lower on the first task benefited more from ChatGPT access. As Fig. 2A shows, the gap between treatment and control is much larger at the left end of the x axis.

Human-machine interactions

What kinds of human-machine interactions underlie the productivity results documented above? Did workers paste the task prompt into ChatGPT and immediately submit its output, minimizing their time spent and increasing their grades because ChatGPT's writing abilities exceeded theirs? Or did they treat ChatGPT as a helpful but imperfect tool, for example, using it to create a rough draft and then spending time editing and improving the draft or using it to brainstorm or edit?

Our evidence supports the first possibility. Almost everyone submitted lightly edited or unedited ChatGPT output, and we observed small time expenditures on editing and no resulting improvement in respondents' grades. In the treatment group, 33% of participants reported submitting ChatGPT's initial output without editing it, and 53% reported editing before submitting. However, those who reported editing were active on the task for only 3.3 min on average after we first observed them pasting in a large quantity of text (presumably from ChatGPT), with most active for 0 to 2 min (*27*). Qualitative examination suggests that most of this editing was superficial, such as changing a placeholder or rearranging a sentence. Evaluator grades also suggest that this editing was ineffectual. There was no correlation between how long a participant was active after pasting in the ChatGPT text and the grade they ultimately received, and treated respondents who used ChatGPT did not receive higher average grades than the raw ChatGPT output that we gave to evaluators to grade (see the supplementary materials).

It is not obvious whether these dynamics should be interpreted as evidence that ChatGPT will displace human workers or evidence that it will augment them. Although ChatGPT directly substituted for participants' effort with little need for human input, it also enabled



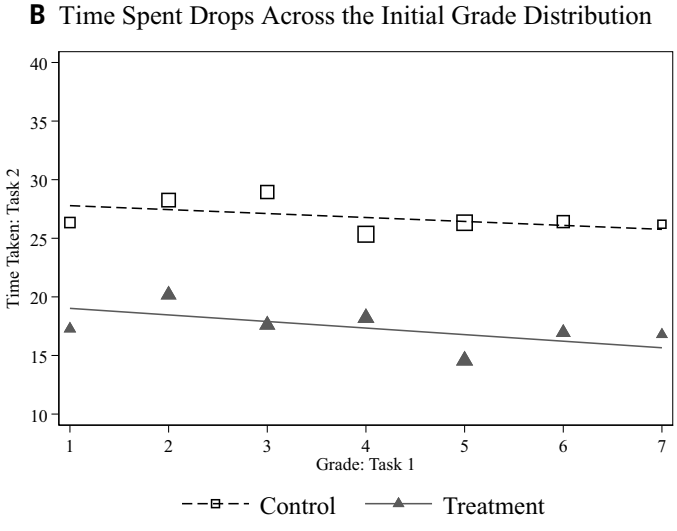


Fig. 2. Effects on grades and time across the initial grade distribution. Participant-evaluator observations are binned together according to the task 1 grade given to this participant by this evaluator. (**A** and **B**) Within each bin, the panels plot the average task 2 grade (A) or task 2 time taken (B) for the observations in the bin separately by treatment versus control. Also shown are the control-group slope, control-treatment difference in slopes, and the 95% confidence interval for the difference. These latter results were calculated from a participant-evaluator level regression of the outcome variable on the task 1 grade, treatment status, treatment*task 1 grade, and grader fixed effects, clustering SEs at the participant level. The control group slope is the coefficient on the task 1 grade, and the difference in slopes is the coefficient on the treatment*task 1 grade interaction. Note that this difference in slopes will not match up exactly with the difference between the raw slopes plotted in the graph because these raw slopes do not use grader fixed effects.

participants to complete tasks much faster. We reflect on this further in the Discussion.

Subjective outcomes: Job satisfaction, self-efficacy, and beliefs about automation

Many of our treated participants had never heard of (30%) or never used (68%) ChatGPT before participating in the experiment. We used a battery of questions to assess their subjective reactions to encountering the technology. As depicted in Fig. 3, participants enjoyed the tasks more by 0.47 SDs when given access to ChatGPT ($P < 0.001$). Treated participants' concern for ($P < 0.01$) and excitement about ($P < 0.001$) future effects of AI on their occupations rose, and their overall optimism increased by 0.2 SDs ($P < 0.05$). These effects disappeared in the 2-week and 2-month follow-up surveys, indicating that they are best interpreted as short-run phenomena reflecting respondents' first experiences with the technology (*28*).

Two-week and 2-month follow-up surveys

One powerful indication of the value of ChatGPT to participants is whether they continued to use it after the experiment, in their actual jobs. To track this, we resurveyed participants 2 weeks and 2 months after their completion of the initial survey, with response rates of 92% and 83%, respectively, and no treatment-control imbalance in response rates.

In the 2-week follow-up, 34% of former treatment group participants reported using ChatGPT in their job in the past week, compared with 18% of control group participants ($P < 0.001$). This large gap in usage fully persisted into the 2-month follow-up, when 42% of the treatment group and 27% of control respondents reported using ChatGPT in their jobs in the past week ($P < 0.01$). The persistence of this gap suggests that the dissemination of ChatGPT into real professional activity is still in very early stages, with usage held back by a lack of knowledge about or experience with the technology.

In the 2-week follow-up, ChatGPT users gave the technology an average usefulness score of 3.66 out of 5.00, somewhat lower than in our main experiment, likely because of the greater length and complexity of real-world tasks. The participants reported using it for a broad range of tasks such as generating recommendation letters for employees, responding to customer service requests, brainstorming, rough-drafting emails, and editing.

Nonusers were divided into three roughly equal-sized groups, reporting that: (i) ChatGPT was not useful in their job, (ii) they did not know about it or did not have an account, or (iii) it was not allowed in their workplace or was usually unavailable during the day. The one-third of nonusers who claimed that it was not useful in their job mostly said that this was because the chatbot lacks context-specific knowledge that forms an important part of their writing. For example, they reported that their writing was "very specifically tailored to [their] customers and involves real time information" or "unique [and] specific to [their] company products."

Discussion

College-educated professionals performing midlevel professional writing tasks substantially increased their productivity when given access to ChatGPT. The generative writing tool increased the output quality of low-ability workers and reduced time spent on tasks for workers of all ability levels. At the aggregate level, ChatGPT reduced inequality. It is already being used by many workers in their real jobs.

These results are consistent with other studies showing productivity-enhancing and equalizing effects of recent AI technologies (*8*, *15*, *16*, *18*). Relative to these studies, we analyzed productivity effects across several occupations and tasks, examined how workers use ChatGPT, measured subjective reactions to the technology, and documented persistent effects of our treatment on ChatGPT usage in real jobs.

Limitations

The experiment had several important limitations. We examined a limited range of occupations and tasks in which ChatGPT may be unusually useful. The tasks demanded clear, persuasive, relatively generic writing, which are arguably ChatGPT's central strengths. They did not require context-specific knowledge or precise factual accuracy. The version of ChatGPT used in this experiment cannot, by its

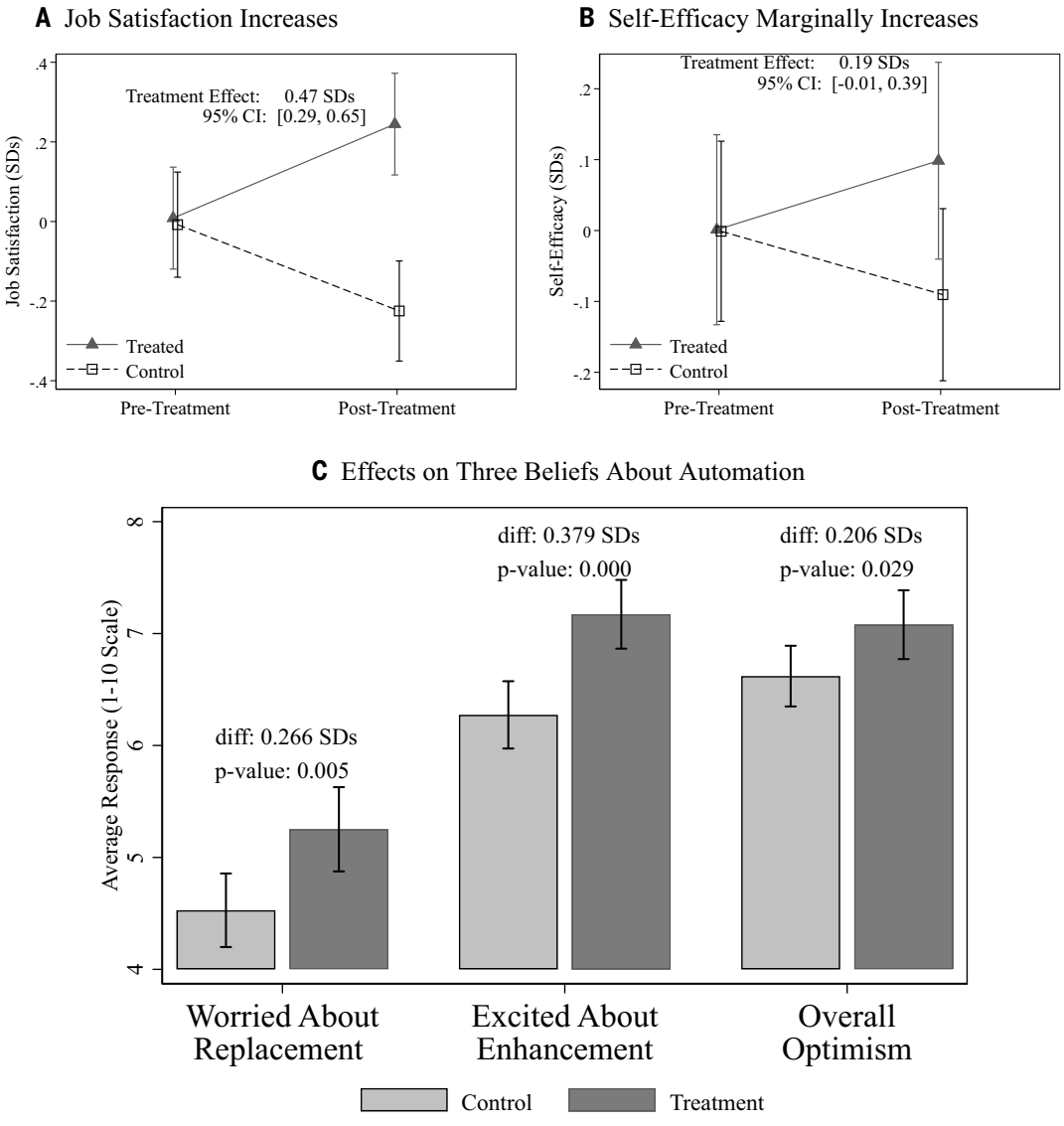


Fig. 3. Job satisfaction, self-efficacy, and beliefs about automation. (**A**) and (**B**) Job satisfaction and self-efficacy (originally elicited on scales of 1 to 10, normalized to have mean = 0 and SD = 1) before and after treatment in the treatment and control groups. Dots are means and error bars are 95% confidence intervals for means. Also shown are the coefficients on the treatment effect of a regression specified as in Fig. 1A. (**C**) Cross-sectional comparison of beliefs about automation in the treatment and control group, all on a scale of 1 to 10. The first question was "How worried are you about workers in your occupation being replaced by AI?" The second was "How optimistic are you that AI may make workers in your occupation more productive?" The third was "How do you feel about the impacts of future advances in AI (where 1 = very pessimistic and 10 = very optimistic)."

nature, access or supply context-specific knowledge and is not a reliable source of precise factual information.

The tasks could also be described through short, self-contained prompts, making use of ChatGPT easy, whereas many real-world tasks involve vaguer objectives and instructions, requiring workers to exercise initiative in determining what to do. Finally, participants in our tasks faced direct incentives in the form of bonus payments scaling with output quality, which encouraged them to maximize generic output quality and minimize time spent. White-collar workers are typically incentivized through longer-run promotion and firing incentives, which might instead encourage conspicuous exertion of effort or the development of a consistent personal style, both of which make ChatGPT less useful.

The tasks and incentive schemes were chosen to meet the constraints of the experimental design. We required short tasks that could be explicitly described for and performed by a range of anonymous workers online, and we needed to incentivize serious effort. Our judgment was that building factual accuracy requirements into the tasks would either result in tasks that felt artificial and unnatural (e.g., requiring participants to Google and report one or two specific facts) or overwhelm our budget for evaluators (e.g., giving participants an open-ended research task and exhaustively fact checking their assertions).

The aforementioned factors limit but do not eliminate the generalizability of our results. In real-world tasks, the need to fact-check ChatGPT's output will reduce its time-saving benefits, but the speed and writing quality increases observed in our experiment are sufficiently large that we suspect that ChatGPT will still often be useful. Moreover, newer versions of ChatGPT are more consistently factually accurate, and some versions can access the internet to fact check themselves. We speculate that in more open-ended, real-world tasks, workers may find iterative rounds of prompting and discussion with ChatGPT useful even if they cannot immediately prompt out a final product. In these contexts, ChatGPT and human workers may be more strongly complementary than in our experiment. The importance of context-specific knowledge will also limit ChatGPT's utility, but there are plausible work-arounds. ChatGPT can be instructed to incorporate lists of context-specific factors, and organizations may be able to build customized ChatGPT-like models. Our follow-up surveys show that many workers do find ChatGPT useful in their real jobs.

Overall, we speculate that, relative to our experimental findings, the direct productivity effects of ChatGPT in the real economy will be somewhat lower and the technology will be

more strongly complementary to human workers. To what extent either of these is true remains an open question.

Implications

Our experiment captured only direct, immediate effects of ChatGPT on worker productivity. We could not examine the complex labor market dynamics that will arise as firms and workers adapt to ChatGPT. Several factors will mediate how the direct productivity impacts of ChatGPT affect wages and employment in exposed occupations. The first is the degree to which demand for goods produced by ChatGPT could expand as ChatGPT-fueled productivity increases make those goods much cheaper. For example, demand for programming services could plausibly expand massively if the price of those services fell. Aggregate programming employment might consequently increase. It is less clear whether demand for advertising or communication could expand as much, potentially entailing a reduction of employment in those sectors as fewer workers are needed to meet the same static demand. As an additional complication, ChatGPT might directly affect the composition of demand. For example, before ChatGPT, a piece of writing signaled that a company had invested at least some human labor, thought, and judgment into a message, which consumers might have appreciated; with this no longer being the case, demand for these messages could decrease (*29*).

The second factor is the nature and scarcity of the human skills best complemented by ChatGPT. Consider, for example, the use of ChatGPT to produce advertising content. Is this best accomplished by one senior advertising manager directly providing high-level guidance to ChatGPT or by 10 junior advertisers carefully designing prompts and editing ChatGPT's output? The answer will determine the structure of employment in the advertising sector. Similarly, suppose ChatGPT is highly complementary to human labor in programming tasks. If ChatGPT's human copilot needs to be an expert programmer capable of directly proofreading its output, then this could raise programmers' wages by boosting their productivity while their expertise remains scarce. If, by contrast, the complementary human role requires only basic programming knowledge and mainly involves checking output and refining natural-language prompts, then the pool of potential programmers would vastly increase and wages could fall even as productivity rises. More generally, tools such as ChatGPT could make expertise more accessible by facilitating learning (*30*).

Finally, the diffusion and effects of ChatGPT will also depend on organizational considerations that our experiment treating isolated individual workers does not address. ChatGPT might interact with traditional promotion and hiring systems based partly on conspicuous exertion of effort. Large language models may be used to monitor or evaluate workers and avoid paying higher wages (*31*). Organizational and societal norms around the acceptability of using tools such as ChatGPT may take time to cohere and may significantly affect adoption of the technology (*32*–*35*).

Overall, the arrival of ChatGPT ushers in an era of vast uncertainty about the economic and labor market effects of AI technologies (*36*–*38*). Our experiment takes the first step toward answering the many questions that have arisen.

REFERENCES AND NOTES

1. D. H. Autor, *J. Econ. Perspect.* **29**, 3–30 (2015).
2. D. H. Autor, D. Dorn, *Am. Econ. Rev.* **103**, 1553–1597 (2013).
3. T. Eloundou, S. Manning, P. Mishkin, D. Rock, arXiv:2303.10130 [econ.GN] (2023).
4. E. W. Felten, M. Raj, R. Seamans, SSRN [Preprint] (2023); http://dx.doi.org/10.2139/ssrn.4375268.
5. M. R. Frank *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 6531–6539 (2019).
6. L. P. Boustan, J. Choi, D. Clingingsmith, "Automation after the assembly line: computerized machine tools, employment and productivity in the United States" (National Bureau of Economic Research, 2022); https://www.nber.org/papers/w30400.
7. D. Acemoglu, P. Restrepo, "Robots and jobs: Evidence from US labor markets" (National Bureau of Economic Research, 2020); https://www.nber.org/papers/w23285.
8. K. Kanazawa, D. Kawaguchi, H. Shigeoka, Y. Watanabe, "AI, skill, and productivity: The case of taxi drivers" (National Bureau of Economic Research, 2022); https://www.nber.org/papers/w30612.
9. R. Arakawa, H. Yakura, M. Goto, CatAlyst: arXiv:2302.05678 [cs.HC] (2023).
10. D. Acemoglu, P. Restrepo, *Am. Econ. Rev.* **108**, 1488–1542 (2018).
11. A. Agrawal, J. S. Gans, A. Goldfarb, *J. Econ. Perspect.* **33**, 31–50 (2019).
12. M. Hoffman, L. B. Kahn, D. Li, *Q. J. Econ.* **133**, 765–800 (2018).
13. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, *Q. J. Econ.* **133**, 237–293 (2018).
14. Productivity gains will translate into higher wages for workers if worker bargaining power or competition between employers for workers is sufficiently high to force employers to share part of the productivity gains with workers, and if the influx of workers into affected occupations is not large enough to completely offset these wage gains (see Discussion section).
15. E. Brynjolfsson, D. Li, L. R. Raymond, "Generative AI at work" (National Bureau of Economic Research, 2023); https://www.nber.org/papers/w31161.
16. S. Peng, E. Kalliamvakou, P. Cihon, M. Demirer, The impact of AI on developer productivity: Evidence from GitHub Copilot. arXiv:2302.06590 [cs.SE] (2023).
17. A. Calderwood, V. Qiu, K. Ilonka Gero, L. B. Chilton, in *Joint Proceedings of the Workshops on Human-AI Co-Creation with Generative Models and User-Aware Conversational Agents co-located with 25th International Conference on Intelligent User Interfaces (IUI 2020), Cagliari, Italy, March 17, 2020*, W. Geyer, Y. Khazaeni, M. Shmueli-Scheuer, Eds. (CEUR, 2018), vol. 2848 of *CEUR Workshop Proceedings*.
18. A. Campero *et al*., arXiv:2206.12390 [cs.HC] (2022).
19. H. Schwabe, F. Castellacci, *PLOS ONE* **15**, e0242929 (2020).
20. A nascent literature has studied applications of machine learning to *predictive* tasks (consisting of yes/no diagnoses) (*11*).
21. E. Peer, D. Rothschild, A. Gordon, Z. Evernden, E. Damer, *Behav. Res. Methods* **54**, 1643–1662 (2022).
22. In each grading session, evaluators graded up to 14 responses. Evaluators received a base payment of $16, plus up to $8 in bonus payments depending on the correlation of their grades with the grades of other evaluators seeing the same responses.
23. This is the average correlation, across every pair of evaluators who saw the same essay, between the grade given by the first evaluator and the grade given by the second evaluator.
24. D. S. Lee, *Rev. Econ. Stud.* **76**, 1071–1102 (2009).
25. The choice to use ChatGPT is uncorrelated with treated respondents' salary, tenure, or grade on the first task.
26. Because of the small sample size in this group, the confidence interval includes zero and the treatment and control group differ in terms of average pretreatment grades (see the supplementary materials).
27. See the supplementary materials for a full description of this analysis.
28. The effects on beliefs about automation may also have dissipated because rising global awareness of the technology led to an equalization of familiarity between the treatment and control groups.
29. Y. Liu, A. Mittal, D. Yang, A. Bruckman, in *CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, New Orleans, LA, 29 April 2022 – 5 May 2022*, S. Barbosa, C. Lampe, C. Appert, D. A. Shamma, S. Drucker, J. Williamson, K. Yatani, Eds. (Association for Computing Machinery, 2022); https://dl.acm.org/doi/10.1145/3491102.3517731.
30. J. Qadir, TechRxiv [Preprint] (2022); https://doi.org/10.36227/techrxiv.21789434.v1.
31. D. Acemoglu, A. F. Newman, *Eur. Econ. Rev.* **46**, 1733–1756 (2002).
32. J. Ayling, A. Chapman, *AI Ethics* **2**, 405–429 (2022).
33. J. Hohenstein *et al*., arXiv:2102.05756 [cs.HC] (2021).
34. M. Suh, E. Youngblom, M. Terry, C. J. Cai, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021*, Y. Kitamura, A. Quigley, K. Isbister, T. Igarashi, P. Bjorn, S. Drucker, Eds. (Association for Computing Machinery, 2021); https://dl.acm.org/doi/10.1145/3411764.3445219.
35. H. Zohny, J. McMillan, M. King, *J. Med. Ethics* **49**, 79–80 (2023).
36. D. Autor, "The labor market impacts of technological change: From unbridled enthusiasm to qualified optimism to vast uncertainty" (National Bureau of Economic Research, 2022); https://www.nber.org/papers/w30074.
37. N. Kshetri, *IT Prof.* **22**, 63–68 (2020).
38. A. Agrawal, J. Gans, A. Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (Univ. of Chicago Press, 2019).
39. Data and code for: S. Noy, W. Zhang, Experimental evidence on the productivity effects of generative artificial intelligence, Open Science Framework (2023); https://osf.io/xd7qw/?view_only=b8dac58dd6f44b979bf81069022fa392DOI10.17605/OSF.IO/XD7QW.

ACKNOWLEDGMENTS

We thank the editor; three anonymous referees; and D. Acemoglu, N. Agarwal, D. Autor, L. Barros, T. Benheim, A. Finkelstein, J. Horton, S. Jaeger, A. Leslie, J. Mejia, I. Noy, L. Noy, E. Partridge, C. Rafkin, A. Rao, N. Roussille, C. Roth, F. Schilbach, B. Schoefer, L. Schubert, A. Shreekumar, V. Vilfort, S. Wu, and participants at the MIT Labor Lunch for helpful comments and conversations. This experiment was approved by the MIT Committee on the Use of Humans as Experimental Subjects (protocol no. 2212000849). **Funding:** This work was supported by an Emergent Ventures grant, the Mercatus Center, George Mason University (S.N.), a George and Obie Shultz Fund grant, the MIT Economics Department (S.N.), and National Science Foundation Graduate Research Fellowship Grant 745302 (W.Z.). **Author contributions:** Conceptualization: S.N., W.Z.; Funding acquisition: S.N., W.Z.; Investigation: S.N., W.Z.; Methodology: S.N., W.Z.; Project administration: S.N., W.Z.; Supervision: S.N., W.Z.; Visualization: S.N., W.Z.; Writing – original draft: S.N., W.Z.; Writing – review and editing: S.N., W.Z. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** This experiment was preregistered at the American Economic Association's registry for randomized controlled trials (https://www.socialscienceregistry.org/trials/10882). All data and code used in the analysis are available at the Open Science Framework (*39*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh2586
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S3
MDAR Reproducibility Checklist

Submitted 20 February 2023; accepted 2 June 2023
10.1126/science.adh2586

QUANTUM OPTICS

Quantum vortices of strongly interacting photons

Lee Drori†, Bankim Chandra Das†, Tomer Danino Zohar, Gal Winer, Eilon Poem, Alexander Poddubny, Ofer Firstenberg*

Vortices are topologically nontrivial defects that generally originate from nonlinear field dynamics. All-optical generation of photonic vortices—phase singularities of the electromagnetic field—requires sufficiently strong nonlinearity that is typically achieved in the classical optics regime. We report on the realization of quantum vortices of photons that result from a strong photon-photon interaction in a quantum nonlinear optical medium. The interaction causes faster phase accumulation for copropagating photons, producing a quantum vortex-antivortex pair within the two-photon wave function. For three photons, the formation of vortex lines and a central vortex ring confirms the existence of a genuine three-photon interaction. The wave function topology, governed by two- and three-photon bound states, imposes a conditional phase shift of π per photon, a potential resource for deterministic quantum logic operations.

Photons do not interact with one another in the optical regime. An effective interaction between photons can be mediated by matter, but, in conventional optical nonlinear media, this interaction is negligible on the level of individual photons. It is only at the ultimate limit of quantum nonlinear optics, realized in specially engineered systems, that a single photon can alter the optical response of the system, rendering an appreciable photon-photon interaction (*1*–*4*).

The realization of strong photon-photon interactions in atomic ensembles can enable deterministic quantum-information processing for optical qubits, such as single-photon transistors and phase gates (*5*–*7*). It can be used to generate nonclassical states of light, such as squeezed, cluster, and repeater states, as a resource for quantum communication, computation, and sensing (*8*, *9*). It can also manifest in interacting quantum gases and fluids of photons, allowing the exploration of exotic many-body physics for light (*10*–*13*).

We realize an extreme regime of quantum nonlinear optics and observe optical quantum vortices generated by the interaction between only two or three photons. Quantum vortices—phase singularities of the wave function—are typically expected in strongly interacting systems of many particles, such as nonlinear optical systems (*14*, *15*), superfluids (*16*–*19*), and hybrid light-matter systems (*20*–*24*). In this work, we induce a strong, effective interaction at the few-photon level by coupling photons to Rydberg atoms in an ultracold rubidium gas (Fig. 1, A and B).

The photons propagate through the medium as light-matter polaritons, at 10^{-6} times the speed of light. Because of the strong van der Waals coupling between the Rydberg atoms, photons that propagate close to each other—at a distance smaller than the so-called blockade radius r_b—experience a local change in the refractive index. For two photons, the optical field can be described by a wave function $\psi(x_1, x_2)$, which is the probability amplitude of having photons at coordinates x_1 and x_2 along the medium (*25*). The change in refractive index at $|x_1 - x_2| \le r_b$ causes the accumulation of excess local phase in $\psi(x_1, x_2)$, which, under the correct conditions, produces vortices in $\psi(x_1, x_2)$.

To understand the vortex formation, we consider a simplified analytical model describing the evolution of $\psi(x_1, x_2)$ in a medium of uniform density and length L using the two-photon Schrödinger equation (*25*, *26*):

$$i\partial_R\psi = -\frac{1}{2m}\partial_r^2\psi + V^{(2)}(r)\psi \qquad (1)$$

Here, $r = x_1 - x_2$ is the relative coordinate of the two photons, and the center-of-mass coordinate along the medium $R = (x_1 + x_2 + L)/2 = 0 \ldots L$ plays the role of time. The incoming, coherent probe field dictates homogeneous initial conditions $\psi(r, R = 0) = 1$. The photon-photon interaction is described by a nearly square potential $V^{(2)}(r) = U/(1 + r^6/r_b^6)$, where U constitutes the change in the refraction index due to the Rydberg blockade, and the (negative) effective mass $m = -U/8$ arises from single-photon dispersion. Because $mU < 0$, Eq. 1 describes attraction between photons (*25*).

A nontrivial prediction of Eq. 1 is the formation of vortex-antivortex pairs in the medium, symmetrically around the potential well (Fig. 1C). Equation 1 has localized and extended eigenstates, denoted as bound and scattering two-photon states (*25*). Considering a single bound state $\psi_{bound}(r)$ with energy E_2, which is the case in our experiment, the initial scattering component is $\psi_{scat} = 1 - \psi_{bound}(r)$. In the crudest approximation, the phase accumulation of the scattering component can be neglected (*27*), and the solution to Eq. 1 reduc

$$\psi(r, R) = e^{-iE_2R}\psi_{bound}(r) + \psi_{scat} \qquad (2)$$

The phase vortices are formed around $\psi = 0$ when the two terms—the bound and scattering components—cancel each other. This result is generic and occurs for a sufficiently long medium and for any number of bound states. Such an interference mechanism for vortex-antivortex pair formation is universal and known from acoustics (*28*) and atomic collisions (*29*) to linear (*30*, *31*) and nonlinear (*32*) optics. However, such vortex-antivortex pair formation has not been observed before in quantum optics because the required strong and prolonged photon-photon interaction had been unattainable.

We quantify the duration and strength of the interaction by the dimensionless parameters $\varphi = UL/2$ and $\lambda = 2|Um|r_b^2$ (*33*), respectively. For vortices to develop, $\varphi\lambda$ should be large. When the interaction is weak, that is, for $\lambda \ll 1$, we find the condition $\varphi\lambda > \varphi_0$, with the threshold phase $\varphi_0 \approx 0.94\pi$. For moderate interactions $\lambda \gtrsim 1$, the condition is $\varphi > \varphi_0$ [see section S2.4 in (*27*)]. These conditions are illustrated by solid lines in the "phase diagram" in Fig. 1D, and they compare reasonably well with the domain of vortex formation calculated numerically for our experimental, nonuniform medium (green shading).

In Rydberg polariton systems, the essential experimental parameters are the total optical depth OD of the medium (e^{-OD} being the on-resonance transmission) and the optical depth of the blockade range $OD_b = OD \cdot r_b/L$. These parameters govern $\varphi \propto OD$ and $\lambda \propto OD_b^2$. In our experiment, we reach $OD \approx 110$ and $OD_b \approx 22$, setting $\varphi \approx 2.7\pi$ and $\lambda \approx 3$ (blue diamond in Fig. 1D). The moderate interaction obtained in previous experiments ($\varphi \approx 1.5\pi$ and $\lambda \ll 1$), albeit being enough to observe the signatures of photonic bound states (*25*, *34*, *35*), were well below the threshold for vortex formation.

Observation of two-photon vortices

Our experiment starts by compressing and trapping an ultracold rubidium cloud with a Gaussian density profile $\rho_0 e^{-x^2/2\sigma^2}$ centered at $x = 0$ and effective length $L = \sqrt{2\pi}\sigma \approx 75\,\mu m$ along the optical axis x. Our maximal peak density $\rho_0 \approx 5 \times 10^{12}$ atoms/cm^3 is a factor of three to six higher than that in previous experiments and is the primary source of our large $\lambda \propto \rho_0^2$. We form Rydberg polaritons using counter-propagating probe and control fields, which together resonantly couple the atomic 5S ground level to the 100S Rydberg level via the 5P level (Fig. 1B). We send, on average, $f \approx 0.25$ probe photons per microsecond.

The probe photons experience the three-level optical susceptibility, except within the

Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 7610001, Israel.
†These authors contributed equally to this work.
*Corresponding author. Email: ofer.firstenberg@weizmann.ac.il

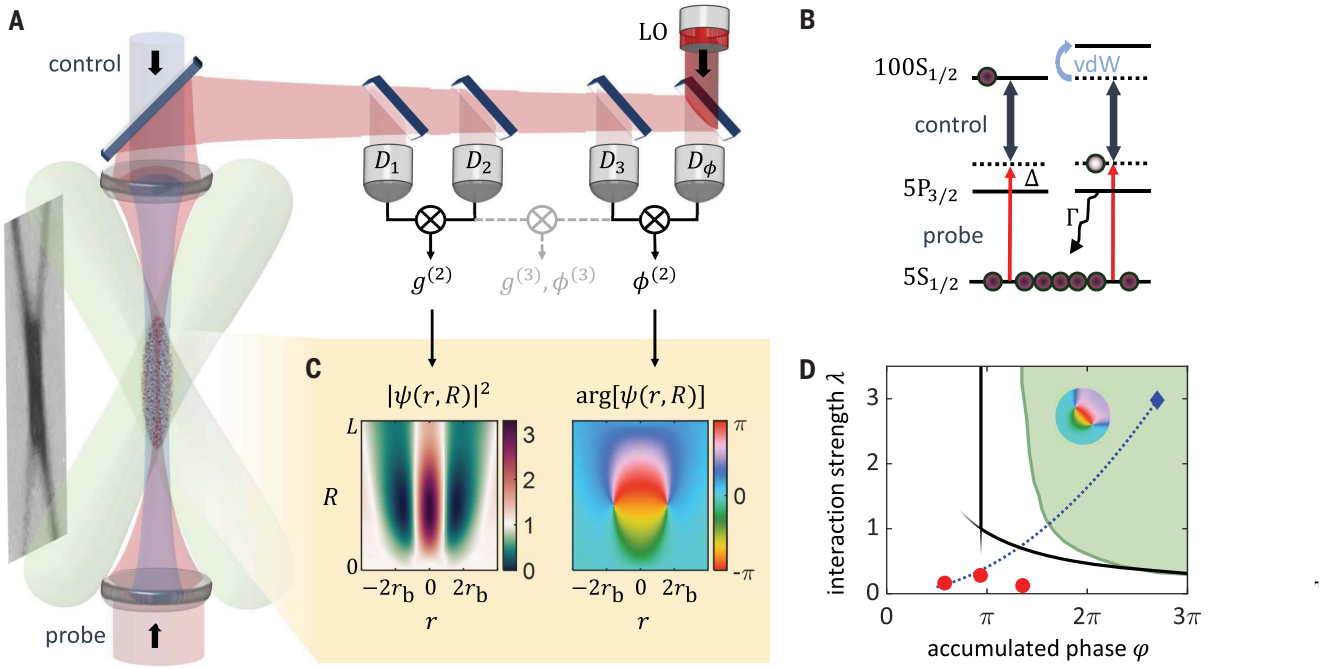


Fig. 1. Setup and conditions for generating photon vortices in Rydberg-mediated quantum nonlinear optics. (**A**) Counter-propagating probe (red) and control (blue) fields are focused onto an elongated ultracold atomic cloud (left, absorption image), held in a crossed optical trap (green). After traversing the medium, the probe light is split and measured by four single-photon detectors D_1 to D_3 and D_ϕ. Detectors D_1 to D_3 provide the two-photon and three-photon correlation functions $g^{(2)}$ and $g^{(3)}$. Detector D_ϕ, which measures the interference beat between the probe and reference light (LO, local oscillator), is correlated with detectors D_1 to D_3 to extract the conditional phases $\phi^{(2)}$ and $\phi^{(3)}$ (see also fig. S2). (**B**) Atomic-level structure for generating polaritons comprising the 100S Rydberg orbital. The van der Waals interaction (vdW) between two Rydberg atoms disturbs the propagation of two close polaritons, leading to local accumulation of excess propagation phase. (**C**) Amplitude and phase of the two-photon wave function, calculated numerically in the Schrödinger approximation (Eq. 1, with $U = \sqrt{12}/r_b$), showing the formation of a vortex-antivortex pair of two interacting photons. (**D**) Conditions for vortex generation (green shaded area, calculated numerically for a realistic Gaussian cloud) in terms of the photon-photon interaction strength λ and the accumulated interaction phase in the finite medium φ. The solid curves are minimal conditions calculated analytically for a uniform cloud under the approximation of Schrödinger evolution (*27*). The dotted blue line indicates the available φ and λ in our experimental setup, peaking at the blue diamond. Red circles indicate the conditions achieved in (*25*, *34*, *35*). The inset image depicts the phase structure of a single vortex-antivortex pair.

blockade range $r_b \approx 15.3$ μm, where they experience the two-level (5S-5P) susceptibility. Ideally, we want the difference U between these susceptibilities to be purely real, rendering a conservative photon-photon interaction. To this end, we first detune the probe from the atomic transition by $\Delta = 9.4\Gamma$, where Γ is the half-width of the 5P level, and then adjust the control frequency such that the transmission remains the same outside and inside of the blockade range (*25*). This adjustment eliminates the residual nonconserving (dissipative) part of the interaction and amplifies the conserving part $U = (\mathrm{OD}/L)(q\Gamma/\Delta)$ by the factor $q \approx 1.4$ (*27*). The atomic density decreases during the experimental cycle by a factor of 5.5, allowing us to study the full range $110 \geq \mathrm{OD} \geq 20$ ($2.7\pi \geq \varphi \geq 0.5\pi$) and, correspondingly, $22 \geq \mathrm{OD_b} \geq 4.1$ ($3 \geq \lambda \geq 0.1$) in each experiment (dotted line in Fig. 1D).

The vortex-antivortex pair forms in the two-photon wave function $\psi(r, R)$ of the probe inside the medium. Although we cannot directly measure these vortices within the medium, we experimentally determine their presence and structure from the correlations of the outgoing photons, essentially measuring $\psi(r, R = L)$ at the medium's boundary. We track the dependence of $\psi(r, R = L)$ on the optical depth OD, and by varying OD, we effectively vary the interaction duration. The dependence on OD at the boundary becomes a proxy for the evolution in the bulk. Whereas OD corresponds to R, the temporal separation τ between the outgoing photons coarsely corresponds to their spatial separation r in the medium (*36*). Experimentally, we measure the two-photon correlation function $g^{(2)}(\tau)$ and conditional phase $\phi^{(2)}(\tau)$ of the outgoing probe field for varying OD and associate them, respectively, with the squared amplitude and phase of $\psi(r, R = L)$.

In the experimental results (Fig. 2, A and B), we first observe the gradual bunching of photons, $g^{(2)}(0) > 1$, as the OD increases, accompanied by depletion of $g^{(2)}(\tau)$ at $|\tau| \approx 0.25$ μs [see cross section (i) in Fig. 2A]. The bunching and depletion are due to the effective attraction between the photons, which is governed by a two-photon bound state and accompanied by accumulation of conditional phase $\phi^{(2)}(0) < 0$ [see (i) in Fig. 2B] (*25*). At $\mathrm{OD} \approx 78$, $\phi^{(2)}(0)$ reaches $-\pi$. A π conditional phase allows for deterministic, maximally entangling operation on photonic qubits (*7*), and here it is obtained for copropagating photons.

The vortex-antivortex pair is formed around $\mathrm{OD} \approx 80$, as manifested by the clockwise and counterclockwise phase twists in Fig. 2B (orange arrows). These phase twists correspond to steep steps of $\phi^{(2)}(\tau)$ at $|\tau| \approx 0.25$ μs, whose direction flips when the OD increases [compare (ii) with (iii) in Fig. 2B]. Furthermore, as expected, the photons are depleted from the vortex core, where $g^{(2)}(\tau) = 0.24$ (global minima in Fig. 2A). Following the vortex-antivortex formation, $\phi^{(2)}(0)$ decreases gradually from π to zero, and the bunching and depletion in $g^{(2)}$ get smaller. The vortices effectively "unwind the tension" between the regions of fast and slow accumulation of phase.

Our experimental results are also corroborated by numerical simulations based on the model described in (*25*, *36*). The simulations use the actual experimental parameters, including the Gaussian density profile of the

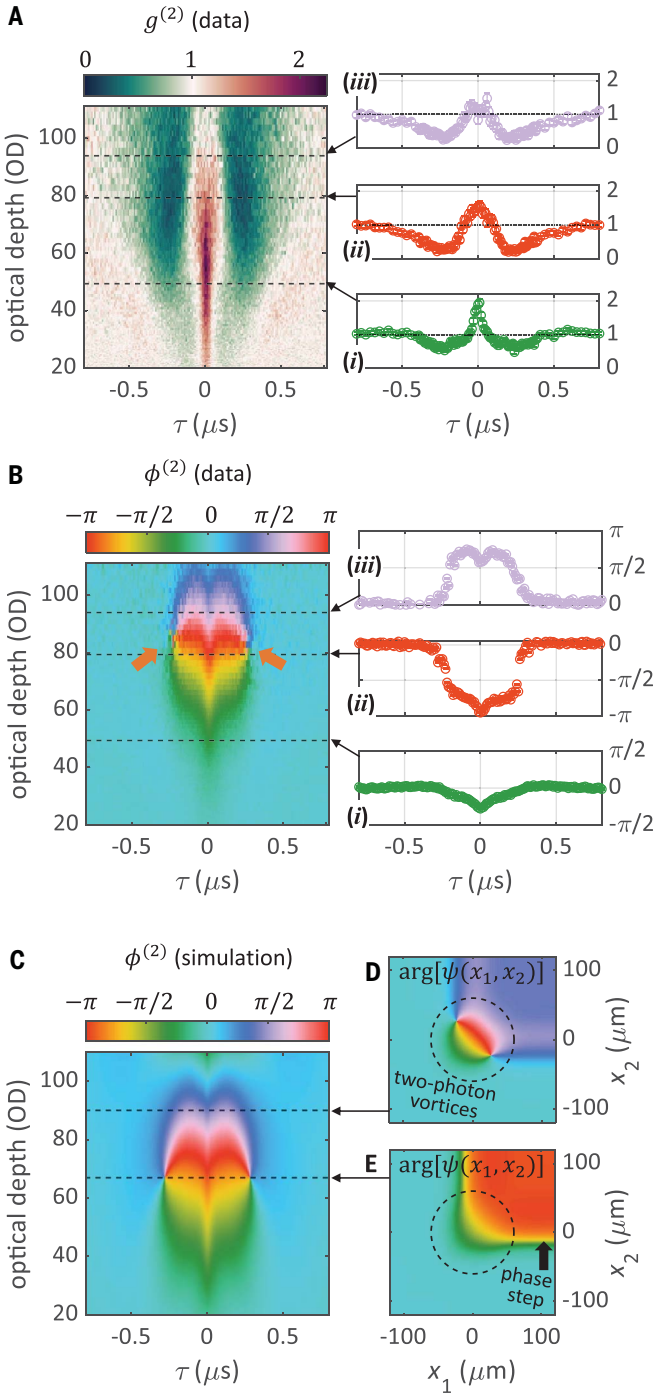


Fig. 2. Two-photon vortices. (**A** and **B**) Color maps show the experimental (A) two-photon correlation $g^{(2)}$ and (B) two-photon phase $\phi^{(2)}$ as a function of the time between photons τ and the optical depth of the atomic medium OD. Enlarged views (i) to (iii) explicitly show the measured curves $g^{(2)}(\tau)$ and $\phi^{(2)}(\tau)$ for OD = 49, 78, and 95. A vortex-antivortex pair [orange arrows in (B)] is formed around OD = 80 on the edge of the medium and is thus captured by the detectors as a phase step around $|\tau|$ = 0.25 μs and a phase wrap of $\phi^{(2)}(\tau = 0)$ from $-\pi$ to $+\pi$. At the vortices' core, $g^{(2)}$ approaches zero. (**C**) Numerically calculated $\phi^{(2)}$ for the experimental conditions. (**D** and **E**) Numerically calculated phase of the stationary, space-dependent two-photon wave function, showing the two-photon vortices forming (D) deep inside the medium and (E) on the edge. Dashed circles represent the edge (2σ) of the Gaussian cloud.

medium and the propagation of the second photon after the detection of the first photon. They are presented in Fig. 2C and are in excellent agreement with the measured data, reproducing the vortex-antivortex pair and the surrounding phase structure in full detail. The 10% discrepancy in the OD at which the vortices are observed is due to a slight saturation of our phase detection scheme and is completely resolved for a weaker incoming probe (see fig. S5). We use these simulations to calculate the conditions for vortex formation in Fig. 1D (green shaded area).

Moreover, the simulations are able not only to reproduce the experimentally accessible data but also to establish the dynamics of the two-photon wave function $\psi(x_1, x_2)$ inside the medium. Figure 2, E and D, presents the phase of $\psi(x_1, x_2)$ for two values of OD. These simulations exhibit the entrance of phase singularities and eventually the presence of a stationary pair of quantum vortices in the medium.

Three-photon vortices

The strong interaction, responsible for forming two-photon vortex-antivortex pairs, leads to an even richer topological structure of the three-photon wave function $\psi(x_1, x_2, x_3)$. We will first illustrate this by considering only the pairwise photon interactions. If only two of the three photons are close together, they attract each other and form a quasi-bound two-photon state. Because the third photon remains unbound, the three pairs of vortices resulting from each of the three possible quasi-bound states would manifest as six vortex lines in the (x_1, x_2, x_3) space.

To verify this prediction, we measured the three-photon phase $\phi^{(3)}(t_1, t_2, t_3)$ and the corresponding correlation function $g^{(3)}(t_1, t_2, t_3)$ over a range of OD. Whereas the stationary two-photon correlation function depends on a single time separation $t_1 - t_2$, characterization of the three-photon wave function requires two time separations. These are conveniently parametrized by the Jacobi coordinates $\eta = t_{21}/\sqrt{2}$ and $\zeta = (t_{13} + t_{23})/\sqrt{6}$, where $t_{ij} = t_i - t_j$ (*37*). The three-photon measurements are then given by three-dimensional (3D) datasets for $\phi^{(3)}$ and $g^{(3)}$ that depend on η, ζ, and OD.

Figure 3A shows $\phi^{(3)}(\eta, \zeta)$ at the critical OD where vortices appear. We identify cores of vortex lines with π-phase dislocations: along the edges of the six-bar star ($|\phi^{(3)}| = \pi$ region) and around the center ($\phi^{(3)} = 0$ region). Figure 3, B and C, shows the development of the phase along the lines $\zeta = 0$ (uniformly separated photons) and $\eta = 0$ (a photon at some distance from a pair). Notably, $\phi^{(3)}(0, 0)$ varies monotonically with OD much faster than $\phi^{(2)}(0)$ does, as seen by comparing Fig. 3B with Fig. 2B.

Eventually, the vortices structure is best understood from Fig. 3D, which shows the reconstructed 3D isosurfaces $|\nabla\phi^{(3)}| = 0.7$ rad/μs, with the phase $\phi^{(3)}$ indicated by the surface color. The starlike structure of tubes in Fig. 3D corresponds to six vortex lines, in agreement with our prediction. We can additionally verify that this structure is due only to the two-photon attraction by constructing it from the measured two-photon data. This construction, shown in fig. S4, exhibits a starlike vortex structure nearly identical to that found in the three-photon data.

The vortex star structure can be explained analytically by generalizing Eq. 2 to three photons. To this end, we replace $\psi_{\text{bound}}(r)$ by a sum $\psi_{\text{bound}}^{(2\to3)} = \psi_{\text{bound}}(r_{21}) + \psi_{\text{bound}}(r_{13}) + \psi_{\text{bound}}(r_{32})$ of three quasi-bound states ($r_{ij} = x_i - x_j$), respecting the bosonic (permutation) symmetry of the three-photon wave function. For simplicity, we assume a weakly bound state $\psi_{\text{bound}}(r) = 2e^{-2|r|/a}$ in a delta potential, where $a \approx r_b/\lambda$ is the scattering length (*25*, *38*). As shown in Fig. 3E using the spatial Jacobi coordinates $\eta = r_{21}/\sqrt{2}$ and $\zeta = (r_{13} + r_{23})/\sqrt{6}$, the generalized Eq. 2 gives the six vortex lines. These results support our

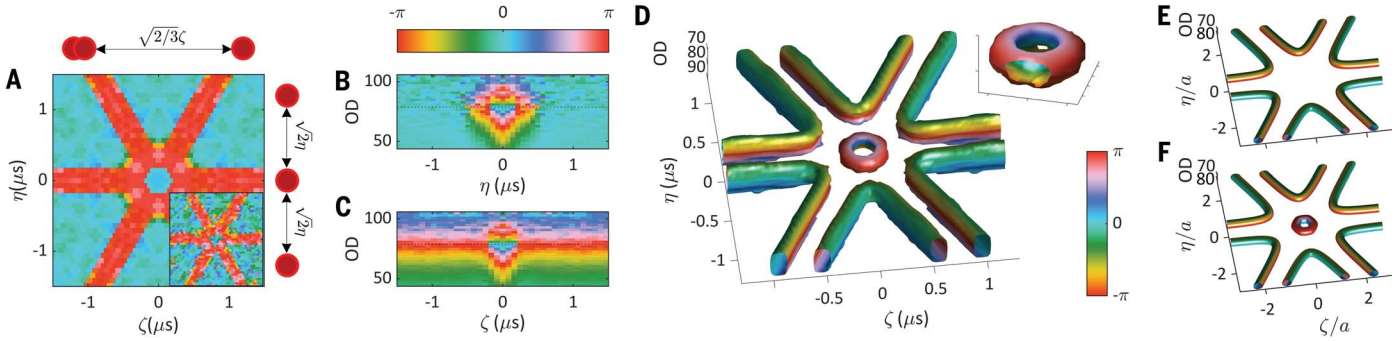


Fig. 3. Three-photon vortex lines and vortex ring. (**A** to **D**) Three-photon conditional phase $\phi^{(3)}(\eta, \zeta)$ measured at different OD, where η and ζ are time separations between the photons in Jacobi coordinates. For each OD, the data are averaged by assuming a sixfold symmetry; an example of data before averaging is shown in the inset in (A). We present three cross sections: (A) OD = 79, (B) ζ = 0, and (C) η = 0. In all these, the central feature originates from three-photon interactions. In (D), to reveal the vortex structure, we plot the measured $\phi^{(3)}$ along isosurfaces of $|\nabla\phi^{(3)}|$, which are themselves derived from the data. We identify six vortex lines (tubes) and a central vortex ring (torus); note the 2π phase twist around the cross section of the tubes and torus (inset). (**E** and **F**) Analytic calculation of the three-photon phase arg[$\psi(\eta, \zeta, R)$] using the approximate ansatz in Eq. 3 (E) without and (F) with the three-photon bound-state term, for $E_3 = 3E_2$ and $\psi^{(3)}_{\text{bound}}(0,0) = -3\psi^{(2\to3)}_{\text{bound}}(0,0)$. The outer vortex lines are due to the generalized two-photon bound states, whereas the vortex ring is due to the three-photon bound state.

interpretation that the vortex lines in the three-photon wave function are a direct generalization of the two-photon vortices.

The full three-photon data, however, show richer behavior, which cannot be reduced to just independent pairwise photon attraction. The effect of a third photon approaching a bound pair is reflected in the central feature in Fig. 3C and, more clearly, as a torus around $\eta = \zeta = 0$ in Fig. 3D, manifesting a vortex ring. This inner ring does not originate from two-photon quasi-bound states but rather from a genuine three-photon bound state $\psi^{(3)}_{\text{bound}}$ (*34*, *37*) that interferes with the scattering states.

To describe the formation of the ring, we use the following ansatz, adding $\psi^{(3)}_{\text{bound}}$ and its energy E_3 to the generalized Eq. 2:

$$\psi(\eta,\zeta,R) = e^{-iE_3R}\psi^{(3)}_{\text{bound}}(\eta,\zeta) + e^{-iE_2R}\psi^{(2\to3)}_{\text{bound}} + \psi_{\text{scat}} \qquad (3)$$

where $\psi_{\text{scat}} = 1 - \psi^{(3)}_{\text{bound}} - \psi^{(2\to3)}_{\text{bound}}$. As shown in (*34*, *39*), the wave function of a three-photon weakly bound state near $\eta = \zeta = 0$ can be approximated by $\psi^{(3)}_{\text{bound}}(\eta,\zeta) \propto e^{-\sqrt{8}|\eta|/a} e^{-\sqrt{2}|\eta-\sqrt{3}\zeta|/a} e^{-\sqrt{2}|\eta+\sqrt{3}\zeta|/a}$. This form can be interpreted as a product of confined states, with three pairs attracting each other simultaneously. It is then straightforward to check, as shown in Fig. 3F, that the interference between the first and last terms in Eq. 3 produces a vortex ring with the same topology that we observe experimentally.

Another important result revealed by the $\phi^{(3)}$ data in Fig. 3 is that the vortex lines and ring appear approximately at the same value of OD ≈ 80. This observation is far from obvious, allowing us to estimate the relation between the bound state energies E_3 and E_2. An analytical model with only pairwise interaction $V^{(2)}$ predicts $E_3 = 4E_2$ (*40*), which would result in the vortex ring appearing at a lower OD. Fitting Eq. 3 to the experimental data yields $E_3 = 3E_2$, so that $e^{-iE_3R} = -1$ forms a vortex ring simultaneously with $e^{-iE_2R} = -1$ forming the vortex lines, as shown in Fig. 3F. The result $|E_3/E_2| < 4$ is experimental evidence of a genuine three-photon repulsion term $V^{(3)}(\eta, \zeta)$ that attenuates the pairwise photon attraction (*37*, *41*). This attenuation follows from the physics of the Rydberg blockade, where one photon can "saturate" the interaction and simultaneously block two (or more) other photons.

The conditional (same-time) phase $\phi^{(3)}_0 \equiv \phi^{(3)}(0,0)$, which originates from the constructive sum of the vortex lines and ring, is $\phi^{(3)}_0 \approx -2\pi$. This nonzero value is evident from the monotonic increase of the phase versus OD in Fig. 3, B and C, or from the phase winding directions in Fig. 3D. Alternatively, one can unwrap the $\phi^{(3)}$ data, as shown in fig. S3, to find -2π at the center of the ring. Importantly, given that $\phi^{(2)}_0 \equiv \phi^{(2)}(0) \approx -\pi$, we find $\phi^{(3)}_0 = 2\phi^{(2)}_0$. This result deviates from the known nonlinear quantum-Kerr phase for n photons $\phi^{(n)}_{\text{Kerr}} = \phi^{(2)}_{\text{Kerr}} n(n-1)/2$, which ascribes $\phi^{(3)}_{\text{Kerr}} = -3\pi$ for $\phi^{(2)}_{\text{Kerr}} = -\pi$ (*42*). Our result $\phi^{(3)}_0 = 2\phi^{(2)}_0$ agrees with the prediction for saturated interaction (*34*), again indicating the attenuation of the interaction by the blockade saturation.

The strong three-photon attraction in the $\lambda > 1$ regime and the repulsive effect of the blockade saturation are also evident in the three-photon correlations $g^{(3)}$ presented in Fig. 4. Whereas the six crests of $g^{(3)} > 1$ away from the center are governed by the two-photon quasi-bound states, the structure at the center is substantially modified by the three-photon bound state. To see this, compare the measured $g^{(3)}$ data in Fig. 4, A to C, with that expected from only the pairwise interactions $g^{(3)}_{\text{d}} = g^{(2)}(t_{21}) + g^{(2)}(t_{13}) + g^{(2)}(t_{32}) - 2$ in Fig. 4, D to F. $g^{(3)}_{\text{d}}$ is the so-called disconnected part of the correlations (*37*). It is evident in all cross sections that the central peak in $g^{(3)}$, reaching a maximal value of $g^{(3)}(0,0) = 7.4 \pm 0.3$, is twice as narrow as the corresponding peak in $g^{(3)}_{\text{d}}$. This strong three-photon bunching has been characterized before and attributed to the tighter confinement of $\psi^{(3)}_{\text{bound}}$ compared with $\psi^{(2)}_{\text{bound}}$ (*34*, *37*).

The effect of blockade saturation is best captured by the connected part of the correlations $g_{\text{c}}^{(3)} = g^{(3)} - g_{\text{d}}^{(3)}$ in Fig. 4, H and I. We can interpret $g_{\text{c}}^{(3)}$ as a manifestation of the contribution of the three-photon correction term $V^{(3)}$, which is repulsive. First, the repulsion attenuates the bunching of simultaneous photons, as evident by the negative $g_{\text{c}}^{(3)}$ close to the center (blue in Fig. 4, G to I). We observe a minimal value of $g^{(3)}_{\text{c,min}} = -2.5 \pm 0.2$. Second, as photons are less attracted to the center, we observe a ring of positive $g_{\text{c}}^{(3)}$ (at a radius $\sqrt{\eta^2+\zeta^2} \approx 0.25$ μs; red in Fig. 4G), reaching $g^{(3)}_{\text{c,max}} = 0.9 \pm 0.1$. As expected, with $\lambda > 1$, both $g^{(3)}_{\text{c,min}}$ and $g^{(3)}_{\text{c,max}}$ are an order of magnitude larger than the values obtained previously with $\lambda \ll 1$ (*34*, *35*, *43*). Finally, $g_{\text{c}}^{(3)}$ is modulated along that ring, exhibiting six peaks where the three photons are uniformly separated in time (e.g., along η, for $\zeta = 0$). This, too, is due to the repulsive correction, which favors equally spaced photons over the asymmetric arrangement of one photon separated from a close pair (e.g., along ζ, for $\eta = 0$). For larger photon separations, the regularization effect of $V^{(3)}$ diminishes, and $g_{\text{c}}^{(3)}$ approaches zero.

Discussion and summary

Our observation of topological defects—including quantum vortex-antivortex pairs, vortex lines, and vortex rings—in the few-body wave function

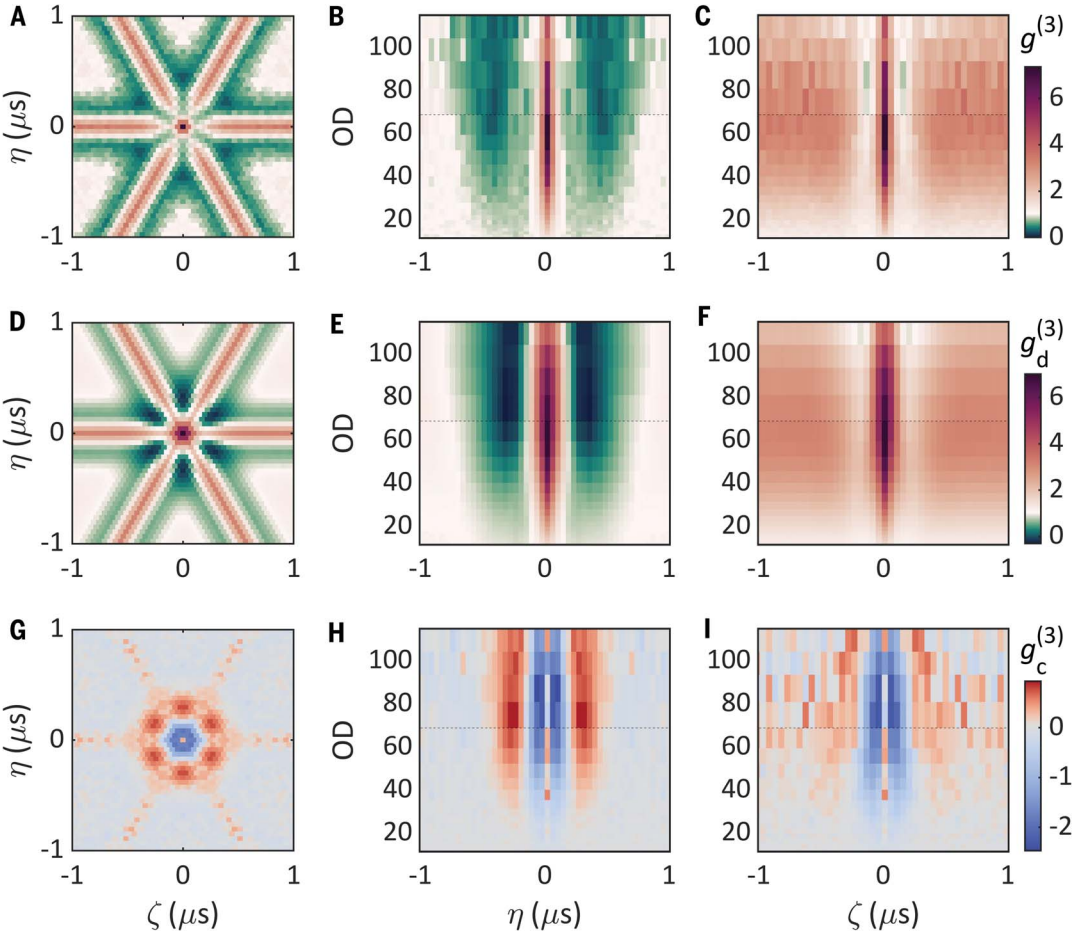


Fig. 4. Three-photon bound-state and blockade saturation. (**A** to **C**) Three-photon correlations $g^{(3)}(\eta, \zeta, \mathrm{OD})$ (after sixfold averaging), sectioned at (A) OD = 69, (B) ζ = 0, and (C) η = 0. (**D** to **F**) Disconnected part of the correlations $g_d^{(3)}(\eta, \zeta, \mathrm{OD})$, calculated from only the $g^{(2)}(\tau, \mathrm{OD})$ data and representing the two-photon contribution to the three-photon correlations. The strong confinement of the three-photon bound state leads to a tighter bunching feature. (**G** to **I**) Connected part of the correlations $g_c^{(3)} = g^{(3)} - g_d^{(3)}$, reflecting the dynamics beyond pairwise interactions. These dynamics are regulated by the saturation of the blockade interaction, expressed as an effective, short-range repulsive force. The negative central feature (blue), the positive ring (red), and the modulation along this ring are all governed by this force.

of interacting photons is enabled predominantly by realizing a strong interaction between the photons. The interaction strength is quantified by the dimensionless parameter λ, which is the ratio between the characteristic interaction and kinetic energies. Whereas in previous works this interaction strength was less than unity, our system reaches $\lambda \approx 3$.

The vortices in the *n*-photon wave function determine the maximal conditional phase $\phi_0^{(n)}$ achievable for copropagating photons. For two photons, the vortex-antivortex pair develops when $\phi_0^{(2)} \approx -\pi$. For three photons, these pairs connect to form six vortex lines surrounding a vortex ring, the sum of which gives $\phi_0^{(3)} \approx -2\pi$. The deviation from the value -3π expected for a quantum-Kerr nonlinearity is a manifestation of the blockade saturation. It demonstrates the role played by the vortices in controlling the conditional phase, with important consequences for deterministic quantum logic operations (*44*, *45*) and other quantum nonlinear devices (*46*, *47*) with three or more photons.

Our findings far from exhaust the potential of exploring many-body polariton physics. Even for relatively weak interactions, one could imagine exotic many-body bound states, such as a cascade of Efimov three-body states with a scaling-invariant spectrum. Although initially proposed for 3D structures (*48*), Efimov states were later predicted to exist for a lower number of spatial dimensions (*49*, *50*), including quasi–1D Rydberg atom clouds (*13*), but have not been observed thus far. Further, our observation of the genuine three-body interaction $V^{(3)}$ could be key to the realization of four-body interactions, which are important for lattice simulations of gauge field theories (*51*).

As opposed to real space, which is limited to three dimensions, the $(x_1, x_2, \ldots)$ space studied in this work can be generalized to more than three dimensions, where higher-dimensional topological structures, such as linked and knotted vortex loops, could be explored. This could provide an interesting connection to the studies of nontrivial topological physics in synthetic dimensions (*52*, *53*). It could further be interesting to examine whether and how the topological phase singularities of the *n*-photon wave function transform to vortices of the mean field in the nonlinear classical optics description at $n \gg 1$.

Finally, a twofold increase in the atomic density will realize the regime $\lambda > \pi^2$, where higher-order two- and three-photon bound states should appear (*26*, *38*). A superposition of vortex series arising from different bound states would drive complex nonperiodic dynamics of the few-photon wave function with an intricate phase structure. The three-photon vortex rings that we observe are a relatively simple example of how the phase of the *n*-photon wave function can be controlled by an additional $(n + 1)$ photon. Potentially, these topological phase structures of *n*-photon wave function could be harnessed to develop new multiphoton control tools and deterministic quantum logic.

REFERENCES AND NOTES

1. A. Imamoğlu, H. Schmidt, G. Woods, M. Deutsch, *Phys. Rev. Lett.* **79**, 1467–1470 (1997).
2. S. E. Harris, L. V. Hau, *Phys. Rev. Lett.* **82**, 4611–4614 (1999).
3. D. E. Chang, V. Vuletić, M. D. Lukin, *Nat. Photonics* **8**, 685–694 (2014).
4. O. Firstenberg, C. S. Adams, S. Hofferberth, *J. Phys. At. Mol. Opt. Phys.* **49**, 152003 (2016).
5. Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, H. J. Kimble, *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
6. H. Gorniaczyk, C. Tresp, J. Schmidt, H. Fedder, S. Hofferberth, *Phys. Rev. Lett.* **113**, 053601 (2014).
7. T. Stolz *et al*., *Phys. Rev. X* **12**, 021035 (2022).
8. J. Hinney *et al*., *Phys. Rev. Lett.* **127**, 123602 (2021).
9. C.-W. Yang *et al*., *Nat. Photonics* **16**, 658–661 (2022).
10. I. Carusotto, C. Ciuti, *Rev. Mod. Phys.* **85**, 299–366 (2013).
11. J. Otterbach, M. Moos, D. Muth, M. Fleischhauer, *Phys. Rev. Lett.* **111**, 113001 (2013).
12. M. F. Maghrebi *et al*., *Phys. Rev. A* **91**, 033838 (2015).
13. M. J. Gullans *et al*., *Phys. Rev. Lett.* **119**, 233601 (2017).
14. V. Tikhonenko, J. Christou, B. Luther-Davies, Y. S. Kivshar, *Opt. Lett.* **21**, 1129–1131 (1996).
15. A. V. Mamaev, M. Saffman, A. A. Zozulya, *Phys. Rev. Lett.* **76**, 2262–2265 (1996).
16. G. Gauthier *et al*., *Science* **364**, 1264–1267 (2019).
17. S. P. Johnstone *et al*., *Science* **364**, 1267–1271 (2019).
18. J. R. Abo-Shaeer, C. Raman, J. M. Vogels, W. Ketterle, *Science* **292**, 476–479 (2001).
19. L. Klaus *et al*., *Nat. Phys.* **18**, 1453–1458 (2022).
20. J. Klaers, J. Schmitt, F. Vewinger, M. Weitz, *Nature* **468**, 545–548 (2010).
21. E. Busley *et al*., *Science* **375**, 1403–1406 (2022).
22. J. Kasprzak *et al*., *Nature* **443**, 409–414 (2006).
23. K. G. Lagoudakis *et al*., *Nat. Phys.* **4**, 706–710 (2008).
24. G. Roumpos *et al*., *Nat. Phys.* **7**, 129–133 (2010).
25. O. Firstenberg *et al*., *Nature* **502**, 71–75 (2013).
26. M. Moos, R. Unanyan, M. Fleischhauer, *Phys. Rev. A* **96**, 023853 (2017).
27. Materials and methods are available as supplementary materials.
28. J. F. Nye, M. V. Berry, F. C. Frank, *Proc. R. Soc. London Ser. A* **336**, 165–190 (1974).
29. J. O. Hirschfelder, K. T. Tang, *J. Chem. Phys.* **64**, 760–785 (1976).
30. K. Y. Bliokh, F. Nori, *Phys. Rev. A* **86**, 033824 (2012).
31. K. Y. Bliokh, *Phys. Rev. Lett.* **126**, 243601 (2021).
32. N. Jhajj *et al*., *Phys. Rev. X* **6**, 031037 (2016).
33. P. Bienias, H. P. Büchler, *New J. Phys.* **18**, 123026 (2016).
34. Q.-Y. Liang *et al*., *Science* **359**, 783–786 (2018).
35. D. P. Ornelas-Huerta *et al*., *Phys. Rev. Lett.* **126**, 173401 (2021).
36. T. Peyronel *et al*., *Nature* **488**, 57–60 (2012).
37. K. Jachymski, P. Bienias, H. P. Büchler, *Phys. Rev. Lett.* **117**, 053601 (2016).
38. P. Bienias *et al*., *Phys. Rev. A* **90**, 053804 (2014).
39. Z. Chen, Y. Zhou, J.-T. Shen, *Opt. Lett.* **45**, 2559–2562 (2020).
40. J. B. McGuire, *J. Math. Phys.* **5**, 622–636 (1964).
41. M. J. Gullans *et al*., *Phys. Rev. Lett.* **117**, 113601 (2016).
42. S. Haroche, J.-M. Raimond, *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
43. N. Stiesdal *et al*., *Phys. Rev. Lett.* **121**, 103601 (2018).
44. X. Wang, A. Sørensen, K. Mølmer, *Phys. Rev. Lett.* **86**, 3907–3910 (2001).
45. C.-P. Yang, Y. Liu, F. Nori, *Phys. Rev. A* **81**, 062323 (2010).
46. M. R. Geller, *Sci. Rep.* **11**, 402 (2021).
47. T. Opatrný *et al*., *Sci. Adv.* **9**, eadf1070 (2023).
48. P. Naidon, S. Endo, *Rep. Prog. Phys.* **80**, 056001 (2017).
49. Y. Nishida, S. Tan, *Few-Body Syst.* **51**, 191–206 (2011).
50. S. Moroz, J. P. D'Incao, D. S. Petrov, *Phys. Rev. Lett.* **115**, 180406 (2015).
51. R. Irmejs, M. C. Banuls, J. I. Cirac, Quantum simulation of Z2 lattice gauge theory with minimal resources. arXiv:2206.08909 [quant-ph] (2022).
52. A. Dutt *et al*., *Science* **367**, 59–64 (2020).
53. A. Balčytis *et al*., *Sci. Adv.* **8**, eabk0468 (2022).
54. L. Drori *et al*., Quantum vortices of strongly interacting photons. Zenodo (2023); https://doi.org/10.5281/zenodo.7962196.

ACKNOWLEDGMENTS

We thank C. Avinadav for helpful discussions. **Funding:** We acknowledge financial support from the Israel Science Foundation, the US-Israel Binational Science Foundation (BSF) and the US National Science Foundation (NSF), European Research Council starting investigator grant QPHOTONICS 678674, the Minerva Foundation with funding from the Federal German Ministry for Education and Research, the Estate of Louise Yasgour, and the Laboratory in Memory of Leon and Blacky Broder. **Author contributions:** L.D., B.C.D., G.W., and O.F. conceptualized the project. L.D., B.C.D., T.D.Z., and G.W. performed the experiment and analyzed the data. B.C.D. and A.P. lead the theory work. All authors contributed to building the methodology and writing the final manuscript. B.C.D., E.P., A.P., and O.F. supervised the project. **Competing interests:** The authors have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials or are deposited at Zenodo (*54*). **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh5315
Materials and Methods
Supplementary Text
Figs. S1 to S6
References

Submitted 9 March 2023; accepted 7 June 2023
10.1126/science.adh5315

MOLECULAR BIOLOGY

Widespread regulatory specificities between transcriptional co-repressors and enhancers in *Drosophila*

Jelle Jacobs[1], Michaela Pagani[1], Christoph Wenzl[1], Alexander Stark[1,2]*

Gene expression is controlled by the precise activation and repression of transcription. Repression is mediated by specialized transcription factors (TFs) that recruit co-repressors (CoRs) to silence transcription, even in the presence of activating cues. However, whether CoRs can dominantly silence all enhancers or display distinct specificities is unclear. In this work, we report that most enhancers in *Drosophila* can be repressed by only a subset of CoRs, and enhancers classified by CoR sensitivity show distinct chromatin features, function, TF motifs, and binding. Distinct TF motifs render enhancers more resistant or sensitive to specific CoRs, as we demonstrate by motif mutagenesis and addition. These CoR-enhancer compatibilities constitute an additional layer of regulatory specificity that allows differential regulation at close genomic distances and is indicative of distinct mechanisms of transcriptional repression.

Animal development and homeostasis critically depend on differential gene expression (*1*, *2*), which is enabled by the precise regulation of transcriptional activation and repression (*3*–*8*). Although repression is often associated with heterochromatin (*9*, *10*), genes can also be silenced in transcriptionally permissive euchromatin (*11*–*14*) by repressive transcription factors (TFs), also termed repressors, that bind to DNA and recruit co-repressors (CoRs). Because CoRs can suppress transcription even in the presence of activators (*15*, *16*), this mode of gene silencing is termed active transcriptional repression (*17*, *18*). Active repression is critical, and its failure can cause developmental defects (*19*–*21*) and diseases such as cancer (*12*, *22*). It is also conceptually intriguing because it requires the overriding of activating cues. However, the modes and mechanisms of this process are unclear.

Given that transcriptional activation can occur by means of distinct modes (*23*–*25*), it is intriguing to speculate whether distinct modes of active repression exist. Examples of specificities between CoRs and activators have indeed been observed. The CoR retinoblastoma protein (Rb) can repress the TFs E2F, Mip120, and PU.1 but not others like SP-1 (*26*–*28*), and CoRs can have different activities in distinct transcriptional contexts (*29*). Comprehensive studies that define different modes of active repression and uncover their regulatory rules are, however, lacking.

In this work, we determined the mutual compatibilities among five known CoRs and a genome-wide library of active enhancers by measuring enhancer-activity changes upon CoR tethering in otherwise unperturbed cells, similar to activator-bypass experiments (*23*, *30*). We reasoned that testing each enhancer with each CoR in all possible combinations should reveal CoR-enhancer combinations that lead to decreased enhancer activity (enhancers are sensitive) and those that do not (enhancers are resistant), indicative of compatible and incompatible pairings.

[1]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Campus-Vienna-Biocenter 1, Vienna, Austria. [2]Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria.
*Corresponding author. Email: stark@starklab.org

Results

UAS-STARR-seq for genome-wide screening of enhancer-CoR specificity

We modified the massively parallel enhancer-activity assay STARR-seq (self-transcribing active regulatory region sequencing) (*31*) to enable the function-based testing of genome-wide enhancer-candidate libraries with different CoRs. Briefly, we introduced four upstream-activating-sequence (UAS) motifs immediately downstream of the enhancers, which leaves the enhancer sequence intact but allows for the direct tethering of selected CoRs through the Gal4 DNA binding domain (Fig. 1A) to assess whether the CoRs can override existing activating cues, a process akin to active repression.

We chose *Drosophila* S2 cells as a model system and five CoRs that represent different protein complexes, repressive pathways, and enzymatic functions (*16*, *32*–*34*): CoRest, CtBP, Rbf, Rbf2, and Sin3A. CoRest, CtBP, Rbf, and Sin3A are conserved, essential, and broadly expressed, whereas Rbf2 has a more restricted expression pattern (*35*, *36*). CoRest mainly represses neuronal genes, directly interacts with histones through its SANT domains, and associates with other conserved repressors, including the demethylase Su(var)3-3 (LSD1) and the histone deacetylase Rdp3 (HDAC1/2) (*32*, *37*). CtBP is recruited by developmental TFs through a short PXDLS (Pro-X-Asp-Leu-Ser, where X is an unspecified residue) peptide motif [illegible] resembles D-2-hydroxyacid dehydrogenases, [illegible] requires nicotinamide adenine dinucleotide (NAD) or its reduced form (NADH) (*38*–*40*). Rbf is a prototypical tumor suppressor of the retinoblastoma family (*41*) and a main transcriptional regulator of cell cycle genes as a part of the DREAM complex, where it counteracts E2F by binding E2F's transactivation domain (*42*–*44*). Rbf2 is a *Drosophila*-specific Rbf paralogue that has specialized and overlapping functions (*34*, *45*). Sin3A is required for repression and activation of target genes of various functions and likely mediates repression through its HDAC-interaction domain (*46*, *47*). Testing diverse CoRs should allow us to detect compatible and incompatible CoR-enhancer pairs. For each CoR, we performed two independent UAS-STARR-seq screens in which we cotransfected the UAS-STARR-seq library with a vector that expresses the Gal4-CoR [or Gal4–GFP (green fluorescent protein) as neutral control; Fig. 1A] and spike-in controls for normalization. In all cases, the two independent replicates correlated well [Pearson correlation coefficient (PCC) > 0.84; fig. S1, A to F]. To reliably assess repression (which requires a high baseline activity), we evaluated enhancer-activity changes for 3094 enhancers that were highly active in Gal4-GFP controls.

CoRs CtBP and Rbf2 silence different subsets of enhancers

We first tested CtBP and found that that some enhancers, such as the enhancers near *Orct*, *Orct2*, and *Pka-C3*, were reproducibly repressed by Gal4-CtBP, whereas others, such as the enhancer near *CG10516*, were unaffected (Fig. 1B). Comparing the two replicates per condition using edgeR showed that CtBP significantly [false discovery rate (FDR) < 0.05, $\log_2$ (fold change) ($\log_2$FC) < 0] repressed 759 enhancers but not the remaining 2335 (Fig. 1C).

Rbf2 was also able to repress only a subset of the enhancers (1733, 56%; Fig. 1D), including the enhancers near *Orct* and *Pka-C3*, which were also repressed by CtBP. In contrast to CtBP, Rbf2 could repress the enhancer near *CG10516* but not the *Orct2* enhancer, indicating that Rbf2 and CtBP have different specificities (Fig. 1B). Indeed, whereas 502 enhancers were repressed by CtBP and Rbf2, 1231 enhancers were repressed by Rbf2 but not by CtBP and 257 enhancers were repressed by CtBP but not by Rbf2 (Fig. 1E).

The enhancer-CoR specificities were validated in luciferase reporter assays that showed that an enhancer near *serpent* was repressed by Gal4-CtBP but not Gal4-Rbf2, an enhancer near CG2116 was repressed by Gal4-Rbf2 but not Gal4-CtBP, and an intronic enhancer of *kay* was repressed by both CoRs, each in agreement with the STARR-seq results (Fig. 1F).

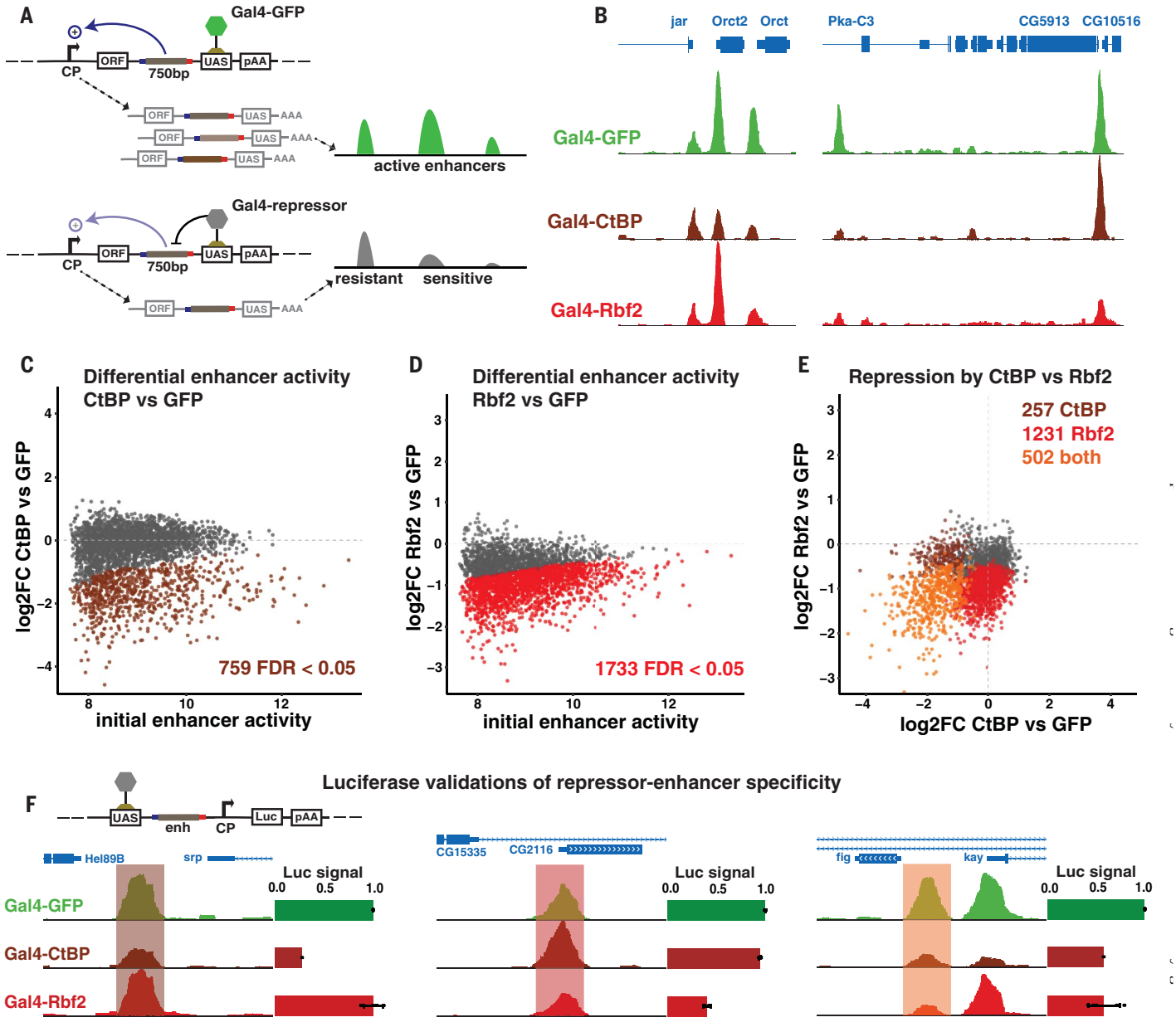


Fig. 1. UAS-STARR-seq maps specificities between CoRs and active enhancers. (**A**) Schematic of the UAS-STARR-seq high-throughput enhancer-CoR assay. Gal4-GFP or a Gal4-CoR is recruited to the UAS sites near a genome-wide library of candidate enhancers. ORF, open reading frame; pAA, poly(A) site. (**B**) Screenshot from the University of California Santa Cruz (UCSC) browser visualizing the activity of nine enhancers (peaks) when GFP (green), the CoR CtBP (brown), and the CoR Rbf2 (red) are recruited. (**C** and **D**) MA plots of initial enhancer activity [$\log_2$(enhancer activity Gal4-GFP)] on the *x* axis and repression by CtBP (C) or Rbf2 (D) on the *y* axis [$\log_2$(enhancer activity Gal4-CoR over Gal4-GFP)]; *n* = 3094. (**E**) Scatterplot contrasting the sensitivity to repression ($\log_2$FC) of each enhancer (*n* = 3094) toward CtBP and Rbf2. Orange indicates enhancers sensitive to both CoRs (FDR < 0.05 and $\log_2$FC < 0), brown indicates enhancers sensitive to CtBP, and red indicates enhancers sensitive to Rbf2. (**F**) Luciferase (Luc) validations. For each plot, the tested enhancer is marked with a transparent red column and the normalized luciferase activity is displayed as horizontal bar plots; *n* ≥ 2.

Differential sensitivity to five CoRs define distinct types of enhancers

Screening three additional CoRs—CoRest, Rbf, and Sin3A—revealed that each was able to repress a specific subset of enhancers (Fig. 2A and fig. S1, G to I). The overall repression profiles of CoRest and Sin3A were similar to that of CtBP (PCCs of 0.83 and 0.7, respectively) and clearly different from Rbf2 (PCCs of 0.28 and 0.24, respectively; fig. S1K); Rbf's repression profile was the most different from that of the other CoRs (PCC < 0.05). The differences between CtBP and Rbf repression persist when they are recruited upstream of the promoter or enhancer, despite changes in CoR efficacies between the positions (supplementary text 1 and fig. S2).

For each CoR, we measured a wide range of repression strengths: Whereas the activity of some enhancers was reduced to background levels, other enhancers were only mildly affected (Fig. 1, C and D, and fig. S1, G to I). Both complete shutdown and mild (or soft) repression are likely biologically relevant (*48*). The CoR-enhancer sensitivities that we observed correlate with where these CoRs operate in the genome, because chromatin immunoprecipitation sequencing (ChIP-seq) signals (*34*, *49*–*51*)

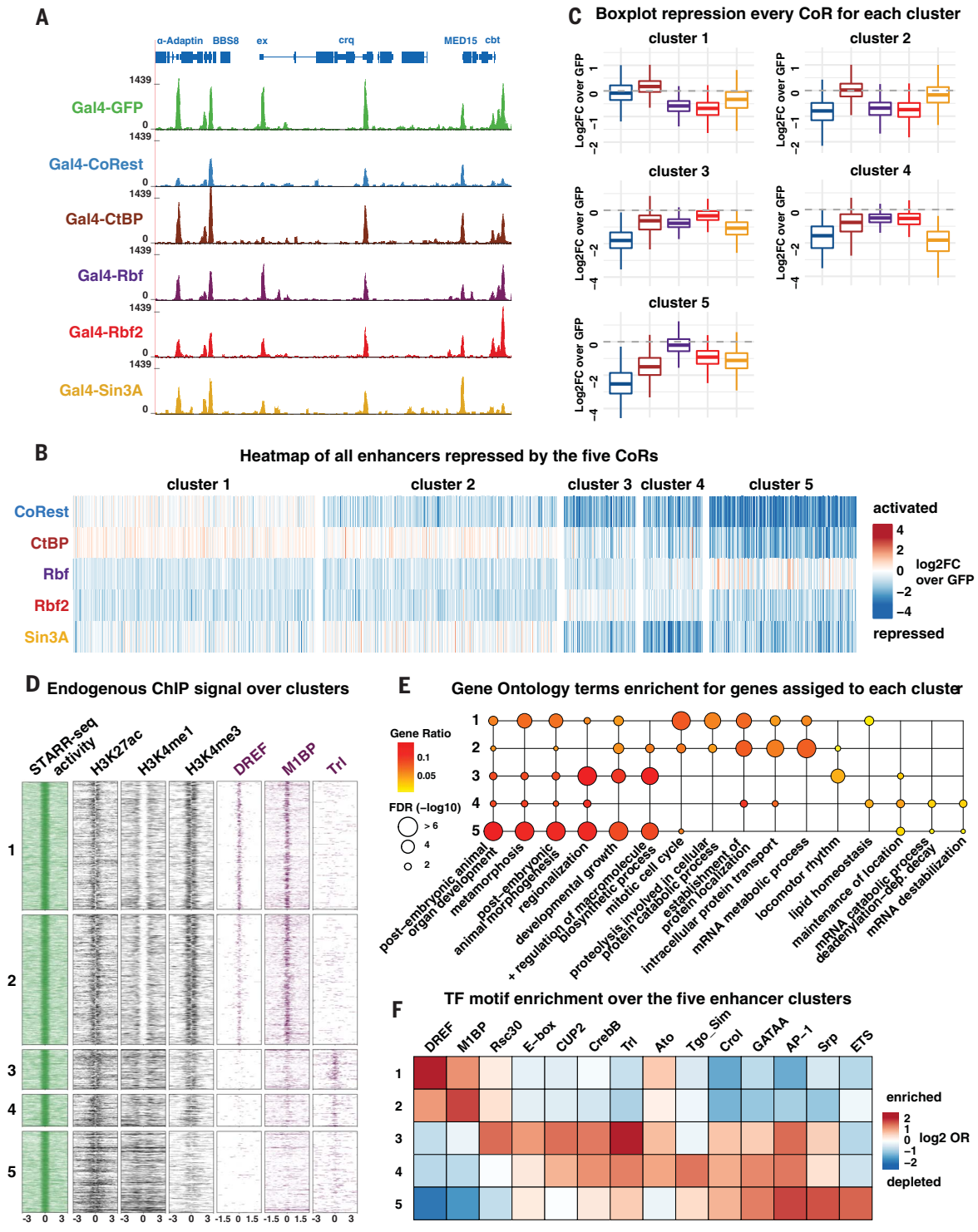


Fig. 2. Enhancer clusters differ in chromatin marks, function, and TF binding. (**A**) Screenshot from the UCSC browser displaying the effect of each CoR on enhancer activity. (**B**) Heatmap visualizing all enhancers from the genome-wide screen, grouped into five clusters based on their similarity in repression profile. The log_2FC in enhancer activity (CoR over GFP) is plotted, and repressed enhancers are marked in blue, unaffected in white, and activated in red. (**C**) Boxplots summarizing the effect that each CoR has on the activity of enhancers from the five clusters. For each boxplot, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values. (**D**) Heatmaps visualizing endogenous ChIP-seq coverage over the enhancers in the clusters. Shown from left to right are STARR-seq activity under Gal4-GFP condition; endogenous chromatin marks H3K27ac, H3K4me1, and H3K4me3; and the TFs DREF, M1BP, and Trl. (**E**) Balloon-plot visualizing GO term enrichment over the five clusters (enhancer to nearest gene). Circle size indicates significance [$-log_{10}$FDR], and the gene ratio is color coded. (**F**) Heatmap of significantly enriched or depleted TF motifs [log_2(odds ratio motifs in selected cluster versus all other clusters)].

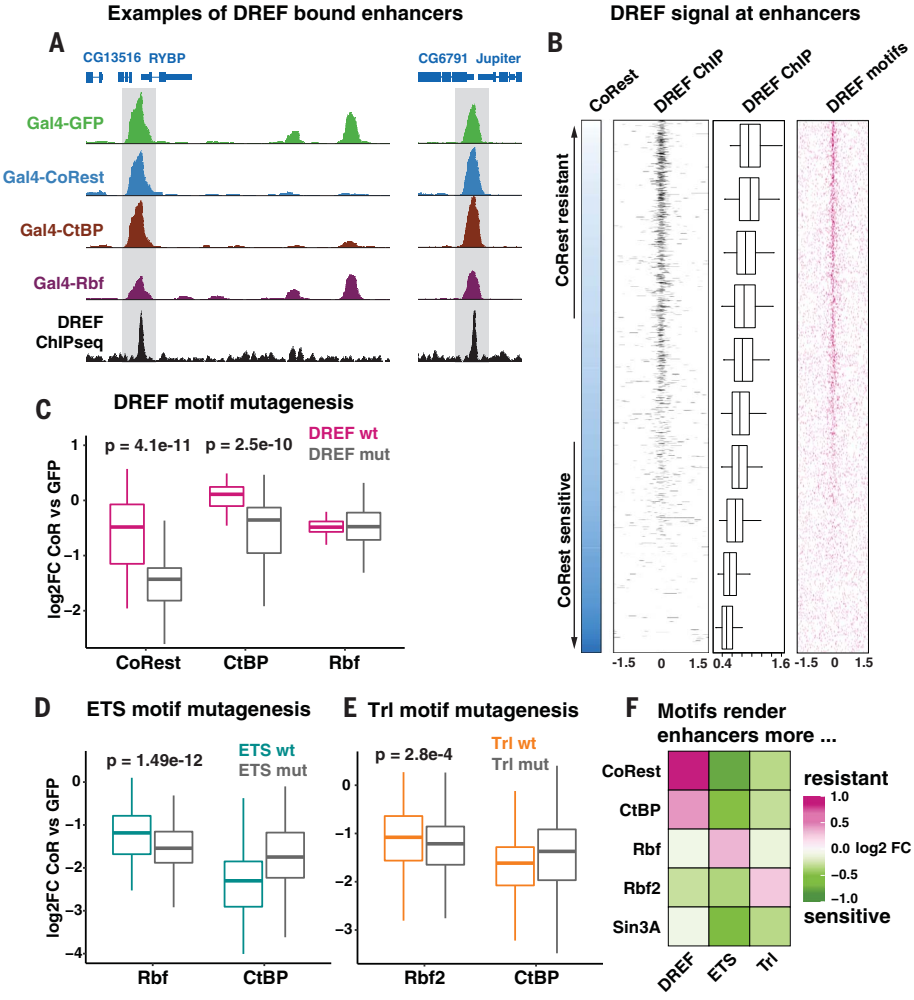


Fig. 3. Specific TF motifs confer resistance against distinct repressors. (**A**) Screenshot from the UCSC browser displaying enhancer activity under GFP, CoRest, CtBP, and Rbf. DREF binding (ChIP-seq) is shown in black. (**B**) Heatmap of all enhancers ranked based on their sensitivity to CoRest. The ChIP-seq signal of DREF is displayed and quantified in boxplots ($p < 2.2 \times 10^{-16}$), and the occurrence of DREF motifs is plotted. (**C**) Boxplot showing the $\log_2$FC of enhancer activity when CoRest (left), CtBP (middle), or Rbf (right) is recruited over GFP. Each enhancer is present as original (magenta) and with its DREF motifs mutated (gray); $n = 65$. mut, mutated; wt, wild type. (**D**) Boxplot showing the $\log_2$FC of enhancer activity when Rbf (left) or CtBP (right) is recruited over GFP. Each enhancer is present as original (cyan) and with its ETS motifs mutated (gray); $n = 157$. (**E**) Boxplot showing the $\log_2$FC of enhancer activity when Rbf2 (left) or CtBP (right) is recruited over GFP. Each enhancer is present as original (orange) and with its Trl motifs mutated (gray); $n = 127$. (**F**) Heatmap summarizing the effects of motif mutations on enhancer sensitivity to the different CoRs ($\log_2$FC of repression wild-type enhancers over repression when specific motif is mutated). For each boxplot, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values.

tend to be enriched at enhancers that the respective CoR is able to repress (fig. S3, A to E). The UAS-STARR-seq results also agree with the differential repression of *PCNA* and *CycB* after overexpressing Rbf or Rbf2 in S2 cells (*45*) (fig. S3, F and G).

We clustered the enhancers into five groups based on their sensitivity to the CoRs ($\log_2$FC against Gal4-GFP) using a self-organizing tree algorithm (SOTA) with the PCC as a distance metric (Fig. 2B). Cluster 1 contains CoRest- and CtBP-resistant enhancers, whereas enhancers in cluster 2 are resistant to CtBP and Sin3A. Cluster 3 contains enhancers that are resistant to Rbf2 and enhancers in cluster 4 are sensitive to all CoRs, whereas enhancers in cluster 5 are overall sensitive to repression but resistant to Rbf (Fig. 2C).

Enhancers clustered by CoR sensitivity differ in chromatin marks and function

We investigated whether these enhancer clusters differ in additional properties that correlate with their behavior toward the CoRs. Initial enhancer activity, as measured by UAS-STARR-seq with Gal4-GFP, was similar in all five clusters (Fig. 2D and fig. S4A). Acetylated histone H3 lysine 27 (H3K27ac), a histone modification that marks active enhancers (*52*), was similarly enriched at the endogenous enhancer loci in all five clusters (Fig. 2D and fig. S4I). This suggests that the distinct specificities did not stem from the differences in initial enhancer strengths. Methylated histone H3 lysine 4 (H3K4me1) and trimethylated histone H3 lysine 4 (H3K4me3) levels did show differential and complementary trends: H3K4me1 showed the highest enrichment in cluster 5, followed by 4 and 3, whereas H3K4me3 was more highly enriched in enhancer clusters 1 and 2. This suggests that clusters 1 and 2 are housekeeping-type enhancers (*24*) (Fig. 2D and fig. S4, B and C), whereas cluster 5 enhancers are developmental or cell type–specific enhancers (*53*). Evaluating chromatin accessibility and H3K27ac levels in other cell types and tissues (*31*, *54*, *55*) confirmed that cluster 1 and 2 enhancers were more globally active, whereas cluster 5 enhancers were S2-cell specific (fig. S4, H to N). The five clusters also differed in the functions of the enhancer-linked genes: Genes linked to Rbf-sensitive clusters 1 and 2 are enriched for gene ontology (GO) terms like mitotic cell cycle or mRNA metabolic process, whereas genes linked to CtBP-sensitive cluster 5 are specifically enriched for development and morphogenesis GO terms (Fig. 2E). These results are consistent with the established functions of Rbf and CtBP in cell cycle regulation and development, respectively (*56*, *57*). Moreover, cluster 5 enhancers were the most strongly repressed by CoRest, CtBP, and Rbf2 (Fig. 2B), which suggests that developmental or cell type–specific enhancers might be more sensitive to repression by these CoRs.

TF binding and motifs differ between enhancer clusters

Because active enhancers are known to function through a variety of TFs, we hypothesized that their differential response to the CoRs might be linked to the specific TFs that they bind. Using published TF ChIP-seq data, we indeed observed that prominent TFs were bound to enhancers from specific clusters and absent from others. The TFs DREF (*58*) and M1BP (*59*), for example, bound almost exclusively to enhancers from clusters 1 and 2, with DREF preferring cluster 1 and M1BP cluster 2 (Fig. 2D and fig. S4, D and E). Conversely, trithorax-like (*60*) (Trl, also called GAGA factor or GAF) was absent from these clusters and instead mainly bound to enhancers from cluster 3 and, to a lower extent, to cluster 4 (Fig. 2D and fig. S4F). The distribution of these three TFs suggests that differential CoR sensitivities might relate to distinct TFs.

To identify associations between CoR sensitivities and TFs in a more comprehensive

manner, we performed TF-motif enrichment analyses for the enhancers in the five clusters using 6502 TF motifs from the iRegulon database (*61*). Consistent with the ChIP-seq results, motifs for DREF, M1BP, and Trl were specifically enriched in clusters 1, 2 and 3, respectively (Fig. 2F). We found additional motifs that were also differentially enriched between the clusters (Fig. 2F). These data together suggest that the different enhancer clusters, defined by their sensitivity toward CoRs, are associated with distinct chromatin features, functions, and TF motif content.

Identification of TF motifs that predict sensitivity and resistance to each CoR

To directly test the association between TF motifs and sensitivity toward each CoR, we evaluated the enrichment of TF motifs within enhancers sensitive or resistant to each CoR (see supplementary text 2 and fig. S5 for a similar analysis of coactivator binding). Motif enrichment profiles across all five CoRs revealed an intricate relationship between enhancer-CoR sensitivity and TF motifs with highly distinctive enrichments (fig. S6, A to F). Consistent with the conserved role of Rb proteins as part of the DREAM complex (*27*, *42*), E2F, DP, and Mip120 motifs were enriched in Rbf-sensitive enhancers. These three motifs, as well as DREF and M1BP motifs, were enriched in enhancers that were resistant to CoRest, CtBP, and Sin3A repression (fig. S6F). Conversely, motifs for the developmental TFs GATA, AP1, and TEAD were enriched in enhancers sensitive to CoRest, CtBP, and Sin3A and resistant to Rbf and Rbf2. Furthermore, Trl and ETS motifs were specifically enriched in enhancers that were resistant to Rbf2 and Rbf, respectively, but sensitive to all other CoRs (fig. S6F). These results indicate that for each CoR, certain TF motifs are specifically and strongly enriched in sensitive and resistant enhancers.

Given the differential motif enrichments, we next tested whether TF motif content is predictive of an enhancer's CoR sensitivity. For this, we trained a generalized linear model (*62*) using TF motif counts as features and enhancer sensitivity to a given CoR as the response variable. For each CoR, when using 10-fold cross-validation, the models based on motif counts performed well and were able to predict an enhancer's sensitivity and resistance to a given CoR (PCCs between 0.46 and 0.73, p values of $<2.2 \times 10^{-16}$; fig. S6G). Overall, these results establish a strong association between CoR sensitivity and TF motif content that is predictive and might correspond to causal relationships.

Distinct TF motifs are required for resistance to specific CoRs

We noticed that DREF motifs were enriched in enhancers that were resistant to CoRest and CtBP (Fig. 2F and fig. S6F). Enhancers bound by DREF, such as enhancers near the *RYBP* and *Jupiter* genes, are resistant to repression by CoRest and CtBP but sensitive to repression by Rbf (Fig. 3A). Ranking all enhancers based on their sensitivity to repression by CoRest or CtBP confirmed that resistant enhancers were significantly ($p < 2.2 \times 10^{-16}$) enriched for DREF motifs and binding according to ChIP-seq (*58*) (Fig. 3B). Given the correlation between CoRest and CtBP resistance and presence of DREF, we considered that DREF might protect enhancers against CoRest- and CtBP-mediated repression. To test whether DREF motifs are required for resistance, we mutated the DREF motifs in 65 enhancers and found that the mutated enhancers were significantly more sensitive to repression by CoRest and CtBP ($p = 4.1 \times 10^{-11}$ and 2.5×10^{-10}), whereas their sensitivity toward Rbf did not change (Fig. 3C and fig. S7, A to C). We conclude that these enhancers require DREF motifs for resistance to CoRest and CtBP.

Several other TF motifs were predicted to confer resistance to repression by specific CoRs. ETS-family motifs, for example, were specifically enriched in enhancers that were resistant to Rbf repression (fig. S6F). We mutated the ETS motifs in 157 enhancers and observed increased sensitivity toward Rbf-mediated repression ($p = 1.49 \times 10^{-12}$; Fig. 3D and fig. S7F). The opposite trend was observed for the other CoRs, which were all able to repress wild-type enhancers containing ETS motifs better than their mutated counterparts, suggesting that ETS TFs are, in general, sensitive to repression (Fig. 3, D and F, and fig. S7, G and I). Similarly, Trl motifs and ChIP signals were enriched in enhancers that were specifically resistant to Rbf2 repression (fig. S8, A and B), and mutating these motifs in 127 enhancers led to a small

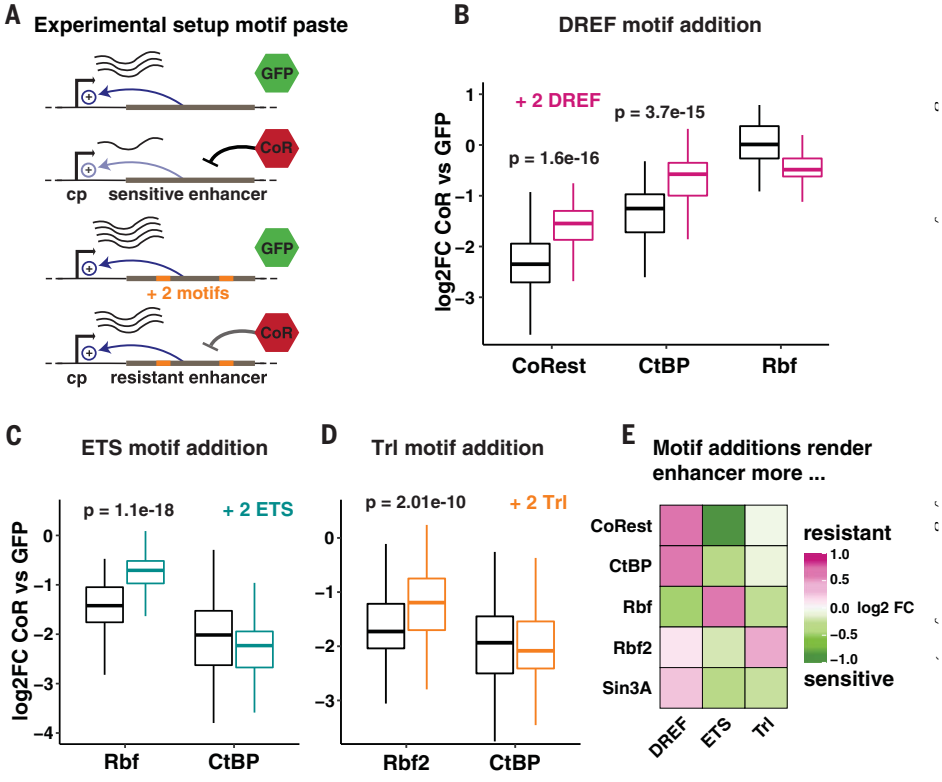


Fig. 4. Effects of TF motif addition on the sensitivity of enhancers to repression. (**A**) Experimental setup of the motif paste experiments. Two TF motifs are pasted into the sequence of sensitive enhancers. The activity of original and motif paste enhancers is measured by UAS-STARR-seq under control (Gal4-GFP) and repressor (Gal4-CoR) conditions (data S2). cp, core promoter. (**B**) Boxplot showing the $\log_2$FC of enhancer activity when CoRest (left), CtBP (middle), or Rbf (right) is recruited over GFP. Each enhancer is present as original (black) and with two DREF motifs pasted (magenta); $n = 110$. (**C**) Boxplot showing the $\log_2$FC of enhancer activity when Rbf (left) or CtBP (right) is recruited over GFP. Each enhancer is present as original (black) and with two ETS motifs pasted (cyan); $n = 107$. (**D**) Boxplot showing the $\log_2$FC of enhancer activity when Rbf2 (left) or CtBP (right) is recruited over GFP. Each enhancer is present as original (black) and with the addition of two Trl motifs (orange); $n = 93$. (**E**) Heatmap summarizing the effects of motif additions on enhancer sensitivity to the different CoRs ($\log_2$FC of repression with motif addition over repression without motif addition). For each boxplot, the center line represents the median, box limits are upper and lower quartiles, and whiskers are minimum and maximum values.

but significant increase in sensitivity toward Rbf2, but not toward the other CoRs (Fig. 3, E and F, and fig. S8, C, D, and F).

Specific TF motifs can turn sensitive enhancers into resistant ones

Given that specific TF motifs are required for CoR resistance, we wondered whether these motifs are also sufficient to drive resistance. To test this, we introduced two "resistant" TF motifs [10 base pairs (bp) each] into CoR-sensitive enhancers at positions deemed unimportant for enhancer activity (*63*) (see table S2) and evaluated the enhancers' sensitivity to repression (Fig. 4A). As a control for the addition of motifs, we also introduced two neutral TF motifs at the same positions (of the ACE2 and FOX TFs), which were predicted to not affect the sensitivity to repression.

Because DREF motifs were required for resistance to CoRest- and CtBP-mediated repression (Fig. 3C), we tested whether they were sufficient to render CoRest- and CtBP-sensitive enhancers more resistant. Overall, the introduction of two DREF motifs was sufficient to protect the enhancers from CoRest- and CtBP-mediated silencing ($p = 1.59 \times 10^{-16}$ and 3.71×10^{-15}, $n = 110$; Fig. 4B and fig. S9, A and B), and this increased protection was specific and due to the DREF motifs, because control motifs had no effect on sensitivity (fig. S9, E and F). Similarly, ETS motifs were necessary (Fig. 3D) and sufficient (Fig. 4C and fig. S10A) to desensitize enhancers from repression by Rbf ($p = 1.06 \times 10^{-18}$, $n = 107$), and Trl motifs were necessary (Fig. 3E) and sufficient (Fig. 4D and fig. S10F) for Rbf2 resistance ($p = 2.01 \times 10^{-10}$, $n = 93$).

Whereas DREF and ETS motifs modulated the enhancers' activities (figs. S7, D and H; S9D; and S10C), Trl motifs had only minor effects on enhancer activity (Gal4-GFP; figs. S8E and S10H). Moreover, in each case, the gain and loss in resistance was CoR specific, because these respective motifs had little effect on the repression by other CoRs or even made the enhancers more sensitive (Fig. 4E and figs. S7 to S10). The addition of ETS or Trl motifs, for example, made enhancers more sensitive to all other CoRs except for Rbf or Rbf2, respectively (Fig. 4E and fig. S10), whereas DREF motifs increased sensitivity toward Rbf ($p = 6.36 \times 10^{-14}$). This might either stem from the CoRs ability to directly antagonize the activity of these additionally bound TFs or because these TFs increase the overall CoR susceptibility of the enhancers. Hence, introducing specific TF motifs changed the enhancers' sensitivity to repression by a given CoR along a spectrum of effect sizes in a specific and predictable manner.

Discussion

In this work, we show that transcriptional enhancers are differentially susceptible to active repression by prominent CoRs. This additional layer of regulatory specificity enables the differential repression of closely spaced enhancers and genes. For example, CoRest could repress the intronic *mys* enhancer without affecting the neighboring *fs(1)h* gene, whereas Rbf could do the opposite (Fig. 5A, left). Similarly, CtBP could repress both enhancers in the *kay/fig* locus, whereas Rbf2 could repress the enhancer closest to *fig* but not the other (Fig. 5A, right).

Although the effects are of varying size, the uncovered specificities between repressors and enhancers reveal a layer of "resistance" against repression (Fig. 5B). Specific motif combinations of activating and repressive TFs, together with the regulatory proteins present in each cell type, will largely determine when and where a regulatory region is active or repressed (*6*, *64*). In addition, the ability of TFs to confer resistance also allows for enhancers and genes to be activated by derepression through TF recruitment, as we demonstrate by motif addition (Fig. 4). This additional layer of regulation circumvents the requirement to remove the involved repressors, and the interplay between TFs and CoRs provides more flexibility to the regulatory system (Fig. 5B).

That one mode of activation can be sensitive to one repressor but resistant to another implies that distinct activation and repression mechanisms must converge. Active repression might directly interfere with specific factors or defined steps of transcriptional activation. Certain CoRs might, for example, inactivate specific TFs or coactivators but not others [e.g., through posttranslational modifications (*65*, *66*)] or counteract their downstream activating mechanisms. Alternatively, certain TFs might directly counteract a repressor's function or bypass the rate-limiting step of initiation or elongation controlled by the CoR. Discerning the distinct mechanisms of activation and repression and how they intersect and coordinate will be of great future interest.

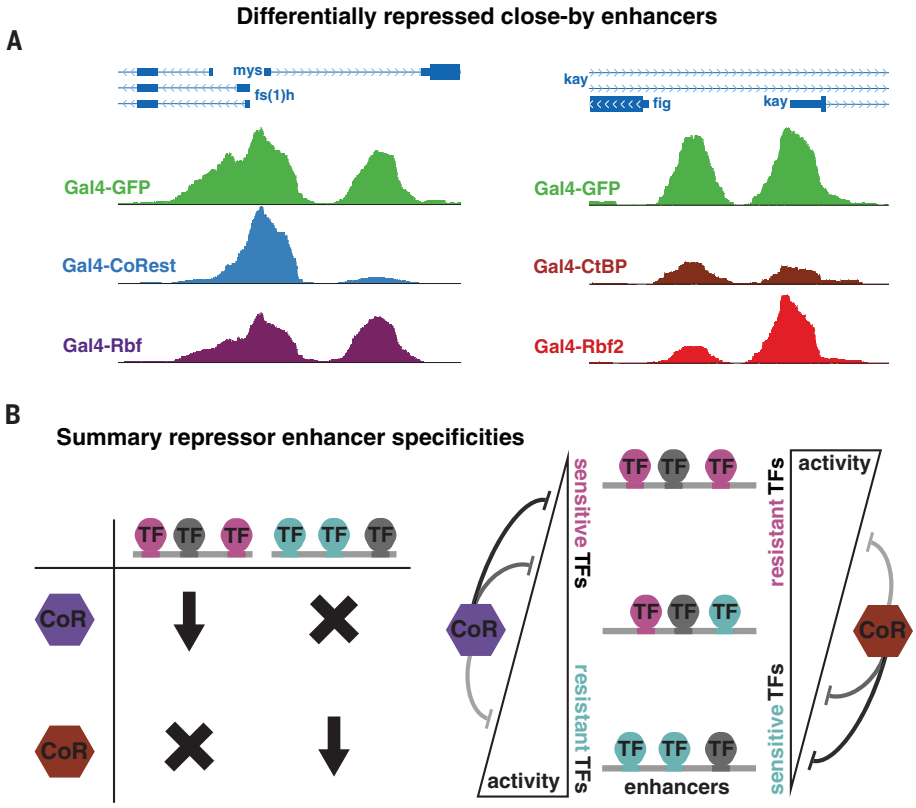


Fig. 5. Implications of repressor-activator specificity. (**A**) Example regions with two closely spaced but differentially repressed regulatory regions. Enhancers differentially repressed by CoRest and Rbf near the *fs(1)h* and *mys* genes are shown on the left, and enhancers differentially repressed by CtBP and Rbf2 near the *fig* and *kay* genes are shown on the right. (**B**) Summary schematics of the uncovered specificities between CoRs and activators, illustrating an additional layer in transcriptional regulation. Depending on motif content (magenta, cyan, or gray), enhancers can be sensitive to one CoR while resistant to another one, and different combinations lead to intermediate responses.

REFERENCES AND NOTES

1. G. M. Cooper, in *The Cell: A Molecular Approach* (Sinauer Associates, ed. 2, 2000).
2. J. H. Gibcus, J. Dekker, *F1000 Biol. Rep.* **4**, 8 (2012).
3. ENCODE Project Consortium, *Science* **306**, 636–640 (2004).

4. D. Shlyueva, G. Stampfel, A. Stark, *Nat. Rev. Genet.* **15**, 272–286 (2014).
5. J. Zuin *et al.*, *Nature* **604**, 571–577 (2022).
6. D. N. Arnosti, S. Barolo, M. Levine, S. Small, *Development* **122**, 205–214 (1996).
7. J. Jaeger, *Cell. Mol. Life Sci.* **68**, 243–274 (2011).
8. S. Barolo, J. W. Posakony, *Genes Dev.* **16**, 1167–1181 (2002).
9. G. Schotta *et al.*, *EMBO J.* **21**, 1121–1131 (2002).
10. M. D. Stewart, J. Li, J. Wong, *Mol. Cell. Biol.* **25**, 2525–2538 (2005).
11. S. E. Keenan *et al.*, *Dev. Cell* **52**, 794–801.e4 (2020).
12. J. Pascual *et al.*, *Dev. Cell* **42**, 667–680.e4 (2017).
13. N. Sharma *et al.*, *Neuron* **102**, 390–406.e9 (2019).
14. J. Zeitlinger *et al.*, *Nat. Genet.* **39**, 1512–1516 (2007).
15. K. Gaston, P.-S. Jayaraman, *Cell. Mol. Life Sci.* **60**, 721–741 (2003).
16. S. Payankaulam, L. M. Li, D. N. Arnosti, *Curr. Biol.* **20**, R764–R771 (2010).
17. W. Hanna-Rose, U. Hansen, *Trends Genet.* **12**, 229–234 (1996).
18. J. B. Jaynes, P. H. O'Farrell, *EMBO J.* **10**, 1427–1433 (1991).
19. R. E. Amir *et al.*, *Nat. Genet.* **23**, 185–188 (1999).
20. J.-Y. Hwang, R. S. Zukin, *Curr. Opin. Neurobiol.* **48**, 193–200 (2018).
21. L. A. Lettice *et al.*, *Dev. Cell* **22**, 459–467 (2012).
22. S. Battaglia, O. Maguire, M. J. Campbell, *Int. J. Cancer* **126**, 2511–2519 (2010).
23. V. Haberle *et al.*, *Nature* **570**, 122–126 (2019).
24. M. A. Zabidi *et al.*, *Nature* **518**, 556–559 (2015).
25. C. C. Bell *et al.*, Transcription factors use a unique combination of cofactors to potentiate different promoter-dependent steps in transcription. bioRxiv 2022.10.25.513774 [Preprint] (2022); https://doi.org/10.1101/2022.10.25.513774.
26. S. J. Weintraub *et al.*, *Nature* **375**, 812–816 (1995).
27. P. Acharya *et al.*, *G3* **2**, 1459–1472 (2012).
28. H. S. Zhang, A. A. Postigo, D. C. Dean, *Cell* **97**, 53–61 (1999).
29. G. Stampfel *et al.*, *Nature* **528**, 147–151 (2015).
30. I. Sadowski, J. Ma, S. Triezenberg, M. Ptashne, *Nature* **335**, 563–564 (1988).
31. C. D. Arnold *et al.*, *Science* **339**, 1074–1077 (2013).
32. M. E. Andrés *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9873–9878 (1999).
33. P. S. Knoepfler, R. N. Eisenman, *Cell* **99**, 447–450 (1999).
34. Y. Wei *et al.*, *G3* **5**, 1503–1515 (2015).
35. O. Stevaux *et al.*, *EMBO J.* **21**, 4927–4937 (2002).
36. S. A. Krause, G. Overend, J. A. T. Dow, D. P. Leader, *Nucleic Acids Res.* **50**, D1010–D1015 (2022).
37. T. Hayakawa, J. Nakayama, *J. Biomed. Biotechnol.* **2011**, 129383 (2011).
38. U. Schaeper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10467–10471 (1995).
39. A.-M. Raicu *et al.*, *Mol. Biol. Evol.* **40**, msad003 (2023).
40. Y. Nibu *et al.*, *EMBO J.* **17**, 7009–7020 (1998).
41. W. Du, M. Vidal, J. E. Xie, N. Dyson, *Genes Dev.* **10**, 1206–1218 (1996).
42. M. Korenjak *et al.*, *Cell* **119**, 181–193 (2004).
43. K. Helin, E. Harlow, A. Fattaey, *Mol. Cell. Biol.* **13**, 6501–6508 (1993).
44. N. J. Dyson, *Genes Dev.* **30**, 1492–1502 (2016).
45. R. Mouawad *et al.*, *Mol. Biol. Evol.* **36**, 2790–2804 (2019).
46. G. E. Adams, A. Chandru, S. M. Cowley, *Biochem. J.* **475**, 3921–3932 (2018).
47. A. Chaubal, L. A. Pile, *Epigenetics Chromatin* **11**, 17 (2018).
48. A. Mitra, A.-M. Raicu, S. L. Hickey, L. A. Pile, D. N. Arnosti, *BioEssays* **43**, e2000231 (2021).
49. M. P. Zappia, A. Rogers, A. B. M. M. K. Islam, M. V. Frolov, *Cell Rep.* **26**, 702–719.e6 (2019).
50. N. Nègre *et al.*, *Nature* **471**, 527–531 (2011).
51. N. Saha, M. Liu, A. Gajan, L. A. Pile, *BMC Genomics* **17**, 111 (2016).
52. M. P. Creyghton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).
53. E. Calo, J. Wysocka, *Mol. Cell* **49**, 825–837 (2013).
54. V. Loubière *et al.*, *Nat. Genet.* **48**, 1436–1442 (2016).
55. J. Jacobs *et al.*, *Nat. Genet.* **50**, 1011–1020 (2018).
56. G. Chinnadurai, *Mol. Cell* **9**, 213–224 (2002).
57. R. A. Weinberg, *Cell* **81**, 323–330 (1995).
58. B. V. Gurudatta, J. Yang, K. Van Bortle, P. G. Donlin-Asp, V. G. Corces, *Cell Cycle* **12**, 1605–1615 (2013).
59. D. G. Baumann, D. S. Gilmour, *Nucleic Acids Res.* **45**, 10481–10491 (2017).
60. N. J. Fuda *et al.*, *PLOS Genet.* **11**, e1005108 (2015).
61. R. Janky *et al.*, *PLOS Comput. Biol.* **10**, e1003731 (2014).
62. J. Friedman, T. Hastie, R. Tibshirani, *J. Stat. Softw.* **33**, 1–22 (2010).
63. B. P. de Almeida, F. Reiter, M. Pagani, A. Stark, *Nat. Genet.* **54**, 613–624 (2022).
64. J. Janssens *et al.*, *Nature* **601**, 630–636 (2022).
65. D. Girdwood *et al.*, *Mol. Cell* **11**, 1043–1054 (2003).
66. K. Higashimoto, P. Kuhn, D. Desai, X. Cheng, W. Xu, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12318–12323 (2007).

ACKNOWLEDGMENTS

We thank the core facilities at the Vienna BioCenter Core Facilities for their excellent support, in particular the molecular biology service, the Next-Generation Sequencing facility, and the High-Performance Computing team. We thank B. P. de Almeida for assisting with the oligonucleotide library design, A. Andersen (Life Science Editors) for comments on the manuscript, and G. Hulselmans and S. Aerts (KU Leuven) for sharing the TF motif position weight matrix collection. **Funding:** This work was funded by European Molecular Biology Organization (EMBO) postdoctoral fellowship ALTF 924-2018 (J.J.) and Marie Skłodowska-Curie Actions postdoctoral fellowship 840729 (J.J.). Basic research at the Research Institute of Molecular Pathology is supported by Boehringer Ingelheim GmbH and the Austrian Research Promotion Agency FFG (A.S.). **Author contributions:** Conceptualization: J.J., A.S.; Methodology: J.J., M.P., C.W.; Investigation: J.J., A.S.; Visualization: J.J.; Funding acquisition: J.J., A.S.; Supervision: A.S., J.J.; Writing – original draft: J.J., A.S.; Writing – review and editing: J.J., A.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All raw and processed next-generation sequencing data related to this paper have been deposited to the Gene Expression Omnibus (GEO) under accession number GSE215143. The UAS-STARR-seq tracks can be viewed on the University of California Santa Cruz (UCSC) browser at https://genome.ucsc.edu/s/jelle.jacobs/rep_STARR_gw_norm_merged. Newly generated UAS-STARR-seq plasmids are available upon request. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse. This research was funded in whole or in part by EMBO (ALTF 924-2018) and the European Union (MSCA 840729), cOAlition S organizations. The author will make the Author Accepted Manuscript (AAM) version available under a CC BY-NC-ND 4.0 international public copyright license.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf6149
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*67–84*)
MDAR Reproducibility Checklist
Data S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

Submitted 2 November 2022; accepted 13 June 2023
10.1126/science.adf6149

QUANTUM OPTICS

Biasing the quantum vacuum to control macroscopic probability distributions

Charles Roques-Carmes[1*†‡], Yannick Salamin[1,2*†], Jamison Sloan[1], Seou Choi[1], Gustavo Velez[1], Ethan Koskas[1], Nicholas Rivera[2,3], Steven E. Kooi[4], John D. Joannopoulos[2,4], Marin Soljačić[1,2]

Quantum field theory suggests that electromagnetic fields naturally fluctuate, and these fluctuations can be harnessed as a source of perfect randomness. Many potential applications of randomness rely on controllable probability distributions. We show that vacuum-level bias fields injected into multistable optical systems enable a controllable source of quantum randomness, and we demonstrated this concept in an optical parametric oscillator (OPO). By injecting bias pulses with less than one photon on average, we controlled the probabilities of the two possible OPO output states. The potential of our approach for sensing sub–photon-level fields was demonstrated by reconstructing the temporal shape of fields below the single-photon level. Our results provide a platform to study quantum dynamics in nonlinear driven-dissipative systems and point toward applications in probabilistic computing and weak field sensing.

One of the most fundamental features of quantum physics are vacuum fluctuations (*1*), which underlie many intriguing phenomena (*2*–*8*). Consequently, there has been interest in measuring vacuum fluctuations, either directly (*9*–*11*), or indirectly (*12*, *13*). Sensitive measurements of energy levels, forces, or fields are often required to discern the impact of vacuum fluctuations. However, certain nonlinear systems are so sensitive to initial conditions that vacuum fluctuations critically influence the outcome of classical observables. Famously, multistable systems that spontaneously break continuous or discrete symmetries reveal the statistical nature of the vacuum (*14*), in which stochastic vacuum forces randomly pick one of the steady states; this makes vacuum fluctuations a source of randomness (*15*). In optics, this phenomenon occurs at the thresholds of lasers and optical parametric oscillators (OPOs) (*16*, *17*). As a result, these optical platforms have been thoroughly explored as generators of true and ultrahigh-rate random numbers (*18*–*23*).

These quantum sources of randomness are also promising for many applications. For example, stochastic components are essential across a range of computing tasks (*24*–*29*). Probabilistic computing (*27*) in particular has shown promise for speeding up optimization and inference tasks (*30*–*33*). Its central building block is a probabilistic bit (*p*-bit): a stochastic logical unit described by a controllable probability distribution (for example, producing 1 with a probability p, and 0 with probability $1 - p$) (*30*). The prospect of leveraging vacuum fluctuations as a source of controllable randomness for photonic probabilistic computing is attractive, given the recent demonstration of ultrahigh-

[1]Research Laboratory of Electronics, MIT, Cambridge, MA, USA. [2]Department of Physics, MIT, Cambridge, MA, USA. [3]Department of Physics, Harvard University, Cambridge, MA, USA. [4]Institute for Soldier Nanotechnologies, MIT, Cambridge, MA, USA.
*Corresponding author. Email: chrc@stanford.edu (C.R.-C.); salamin@mit.edu (Y.S.)
†These authors contributed equally to this work.
‡Present address: E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA.

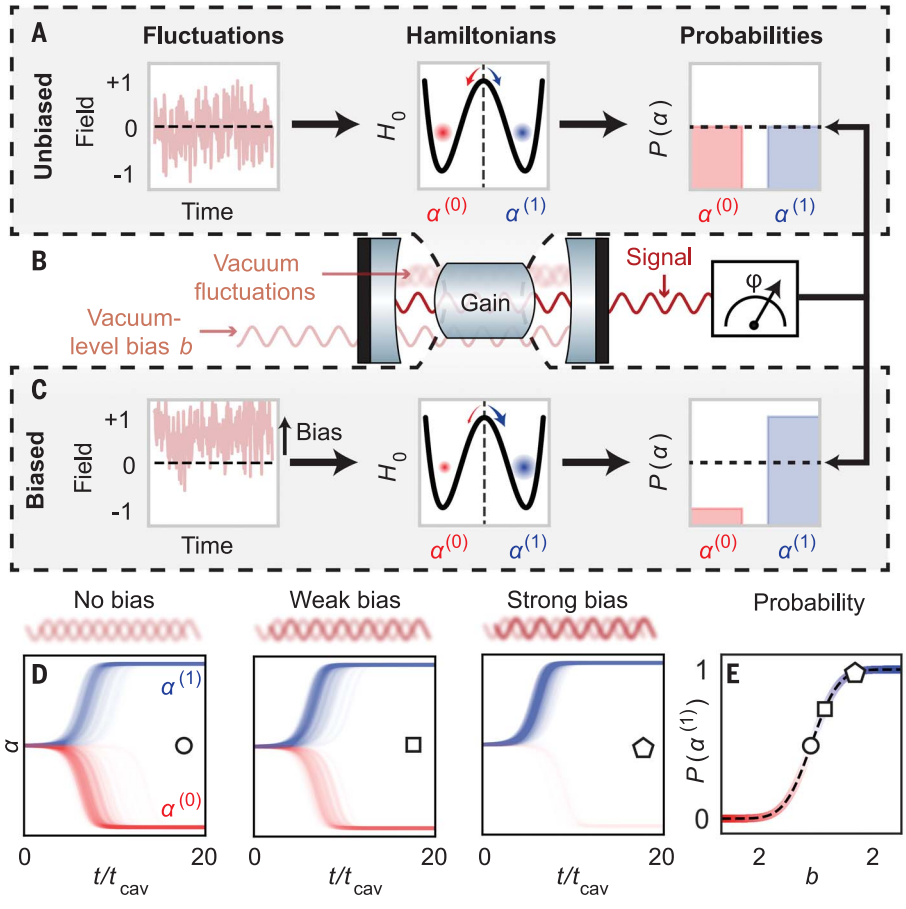


Fig. 1. Tuning macroscopic probability distributions by biasing vacuum fluctuations. (**A** and **C**) Fluctuation snapshot, Hamiltonian, and associated probabilities for an (A) unbiased and (C) biased bistable system. In (C), a vacuum-level bias field is injected in the cavity. (**B**) Schematic of quantum optical multistable system. (**D**) Stochastic amplitude trajectories for a biased optical parametric oscillator (OPO). One hundred trajectories are plotted in each graph, and time is normalized to the cavity lifetime t_{cav}. (**E**) Probability of state $\alpha^{(1)}$ as a function of bias intensity b.

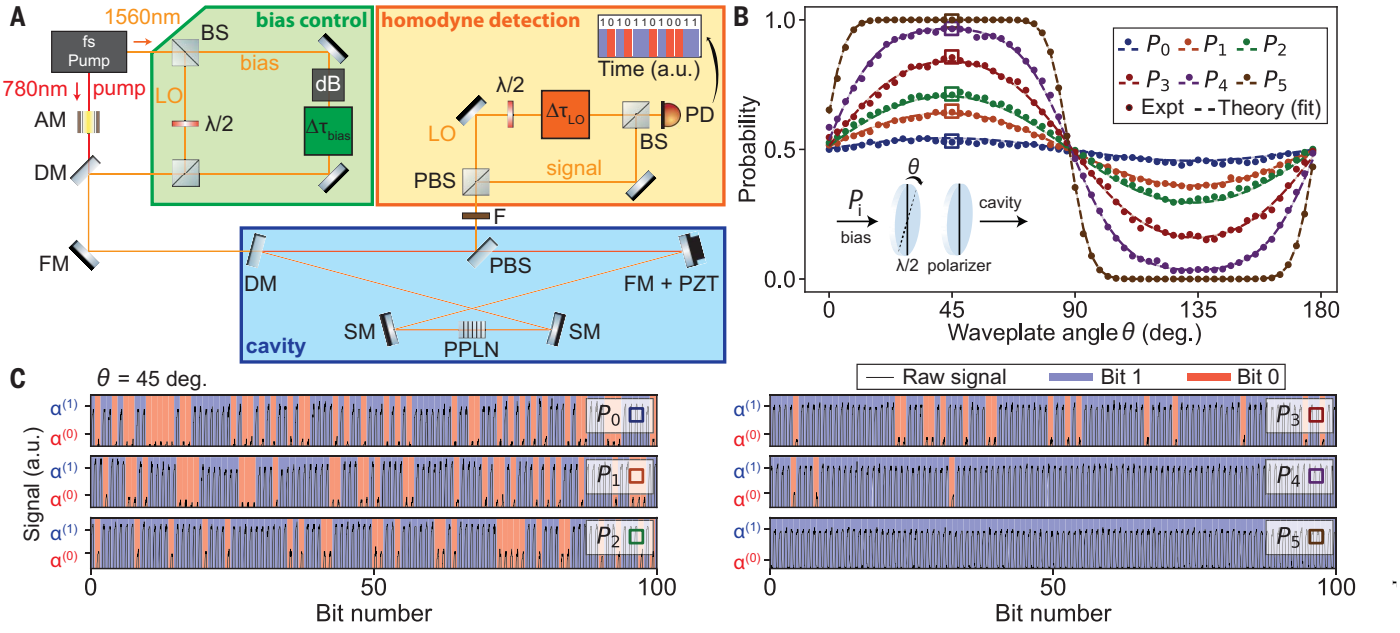


Fig. 2. Experimental demonstration of a photonic *p*-bit in a biased OPO. (**A**) Schematic of the experimental setup. AM, amplitude modulator; BS, beam splitter; PBS, polarization beam splitter; dB, bias attenuation (combination of neutral density filters, pinhole, and polarization optics); λ/2, half waveplate; DM, dichroic mirror; PD, photodiode; FM, flat mirror; SM, spherical mirror; PPLN, periodically poled lithium niobate nonlinear crystal; LO, local oscillator; F, spectral filter; and PZT, piezoelectric actuator. (**B**) Probability measurement for various maximum bias powers P_i and varying waveplate angle θ, over 20,000 samples. (Inset) Schematic of the bias control setup in this experiment. (**C**) Bit stream portions at θ = 45°.

bandwidth random-number generators (*23*) and their potential integration in low-latency and power-efficient photonic processors (*34*).

A substantial gap still separates the current state of the art from the proposed controlled quantum randomness in optics. Specifically, all existing optical random-bit generators rely on perfect symmetry, so that vacuum fluctuations yield unbiased outcomes. The other extreme for nonlinear optical systems lies in externally seeded devices (*35*–*37*), in which the steady state is completely determined, sacrificing any notion of randomness. The existence of these two extremes—perfect unbiased randomness and complete determinism—raises questions about what physics may lie in between.

In this work, we show that the injection of vacuum-level fields into a multistable optical system enables a controllable source of biased quantum randomness. We show that quantum biased randomness can be realized in various nonlinear photonic systems, and we demonstrate our concept in a biased OPO, where the random variable is the phase of the signal field, which can take on values 0 and π. We generated a photonic *p*-bit, with a parameter p that can be controlled by the bias amplitude and phase. We then elucidated the physical origin of this biased symmetry breaking by showing that this biased randomness arises directly from interference between the bias field and vacuum fluctuations. Last, we show that vacuum-level biased optical systems can probe vacuum field fluctuations, as well as the temporal profile of very weak electromagnetic pulses, down to much less than one (≈10^{-3}) photon per pulse.

Controlling macroscopic observables with coherent injection of vacuum-level fields in quantum optical multistable systems

We considered a physical system with Hamiltonian H_0, which possesses multiple degenerate stable steady states (multistable). For concreteness, we considered a bistable system described by two degenerate ground states $\alpha^{(i)}$ ($i \in \{0, 1\}$) (Fig. 1A). The system is initiated in the unstable state that lies at the local maximum of the potential between the two degenerate states. This unstable initial state will then decay into one of the degenerate energy minima with equal probability (Fig. 1A, center). The resulting random outcome stems from random field fluctuations (Fig. 1A, left) in the initial state. Nonlinear quantum optical multistable systems provide a natural platform to implement H_0, and macroscopic probability distributions can be extracted by repeated phase measurements (Fig. 1B).

We focused on one possible implementation of Fig. 1B: a bistable optical system based on a degenerate biased OPO. The essential component of a degenerate OPO is a second-order nonlinear crystal in an optical cavity. When a pump with frequency 2ω is coupled into the crystal, oscillations at the signal frequency ω are induced through parametric amplification. The resulting steady-state signal field can take on two distinct phases, 0 and π—respectively, $\alpha^{(1)}$ and $\alpha^{(0)}$. In the absence of any injected bias field at the signal frequency, the OPO steady state above threshold approaches a mixed state of the two possible phases, with equal likelihood (*20*).

The situation changes when a bias field is injected into the cavity (Fig. 1C). Although adding a coherent field is known to break the symmetry of such systems, we specifically considered vacuum-level bias fields, which are on the order of the fields associated with vacuum fluctuations. By tuning the injected bias field, a mixed-state density matrix is created of the form $\rho = (1-p)|\alpha^{(0)}\rangle\langle\alpha^{(0)}| + p|\alpha^{(1)}\rangle\langle\alpha^{(1)}|$, realizing a binomial distribution $\mathcal{B}_p$: $P(\alpha^{(0)}) = 1 - p$ and $P(\alpha^{(1)}) = p$, with $p \in [0, 1]$. The probability p is given by the analytical expression

$$p(b) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{|b|\cos\phi}{b_0}\right) + 1\right] \qquad (1)$$

In this expression, erf is the error function; b is the bias field, which is defined so that b^2 is the number of bias photons present in the signal pulse in the cavity during each round trip; and ϕ the phase of the bias field relative to the OPO signal. Furthermore, b_0 is the critical bias field level required to modify the probability, which is of order unity, corresponding to single-photon–level bias. Amplitude trajectories for an OPO with weak gain and various levels of bias power are shown in Fig. 1D.

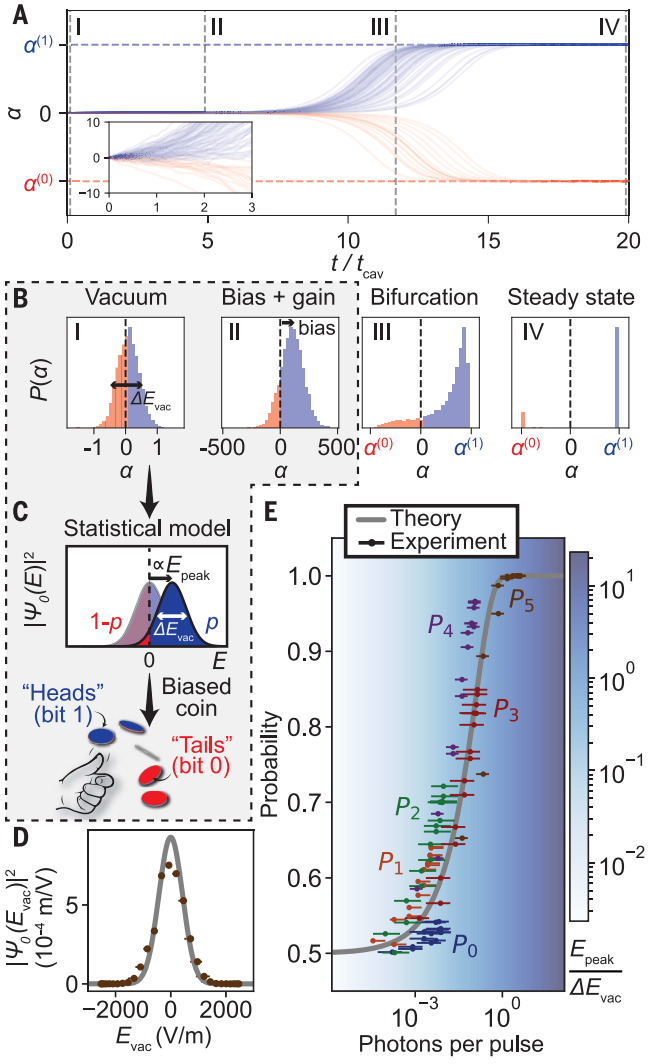


Fig. 3. Biased coin flips from the quantum vacuum. (**A**) Stochastic trajectories of OPO amplitude. (Inset) Zoom-in of the trajectories at early times. (**B**) Four-step model evolution of the probability distribution. (**C**) Statistical model of the biased OPO as a biased coin flip, where the parameter p is determined by the positive-value area under the ground-state wave function density $|\Psi_0(E)|^2$ shifted by E_{peak}. (**D**) Reconstructed wave function density $|\Psi_0(E)|^2$ of the electromagnetic ground state. (**E**) Transition between purely random and deterministic regimes of the OPO. The gray line is the theoretical prediction without any fitting parameters. For readability, only every other angle data point with probability $p > 0.5$ is shown.

The derivation of the bias-probability relationship from a rigorous quantum mechanical model is given in the supplementary text, section B (*38*). The theory consists of writing the Heisenberg-Langevin equation of motion for the weakly biased OPO system and solving an equivalent set of classical stochastic differential equations. A generalization of this concept to multistable systems (such as lasers) and considerations on the implementation of p-bits in ultrafast optics are presented in the supplementary text, section A (*38*).

Experimental demonstration of a photonic p-bit in a biased degenerate optical parametric oscillator

We then demonstrated a vacuum-level biased OPO (Fig. 2A) that consists of three main parts: (i) a degenerate OPO in a bow-tie cavity (Fig. 2A, blue shaded area); (ii) OPO phase interferometric measurement (Fig. 2A, orange shaded area); and (iii) temporal and amplitude control of the bias field (Fig. 2A, green shaded area). A complete description of the experimental setup is given in (*38*) and the supplementary text, section C.

The measured probability p is shown in Fig. 2B, where the bias field amplitude sets the p-bit probability. The bias field amplitude was set by a fixed attenuation (corresponding to the maximum power P_i) and fine tuned by rotating the half waveplate angle, so that the bias power is given by $P_i \sin^2(2\theta)$. We fitted our experimental data to the functional form from Eq. 1, showing excellent agreement with our model. The bit probabilities change in response to bias field attenuation as expected from theory.

By changing the attenuation level of the bias field, we were able to traverse the continuous space between perfect randomness and determinism. Segments of the bit stream statistics (Fig. 2C) show a transition from a slightly skewed binomial distribution $\mathcal{B}_p$ with $p \approx 0.5$ (P_0) to purely deterministic outcomes (P_5). The realized p-bit exhibits the main properties necessary for its use in probabilistic computing platforms (*27*): statistical independence of consecutive samples and continuous tunability of the probability distribution (supplementary text, section A). Because the p-bit outcome is in the optical domain, it could be further processed by an optical analog processor (*34*).

Sensing vacuum fluctuations and sub–photon-level fields with photonic p-bits

We next demonstrated that the observed probability control is the result of bias fields at the level of vacuum fluctuations. Shown in Fig. 3A are sample trajectories simulated by using stochastic differential equations derived from a rigorous quantum mechanical model of a biased OPO. Examining the probability distribution of fields at four characteristic times (marked I to IV) during evolution from the initial condition allowed us to develop a four-step model of how the biased distribution develops, from the initial vacuum fluctuations (I), undergoing bias and gain (II), then bifurcating (III), before reaching steady state (IV).

Our experimental data are quantitatively consistent with these theoretical predictions. The experiments shown in Fig. 2B are realized with $\approx 3 \times 10^{-5}$ to 4 photons per pulse (depending on the total attenuation level). In Fig. 3E, we plotted the measured probability from various bias power levels P_i against the estimated number of photons per pulse, matching our theory without any fitting parameter. Our theory accurately predicted the absolute level of bias fields required to influence the probability in our experiments (on the order of one photon per pulse).

Equivalently, the critical bias peak field is comparable with the amplitude of the vacuum fluctuations ΔE_{vac} of the bias mode. As shown in stage II of Fig. 3B, early stages of the OPO evolution play a critical role in picking the steady state of weakly biased OPOs. This observation allowed us to extrapolate a simple statistical model of the phase measurement process in a biased OPO (supplementary text, section D), where the probability parameter p is given by the area under the positive-field side of the ground-state wave function density $|\Psi_0(E_{vac})|^2$ displaced by E_{peak} (Fig. 3C).

This simple model, which is in agreement with our theory, also confirms the origin of the randomness in our experiment and all other OPO-based random-number generators: quantum statistical fluctuations of the fields in the early stages of field amplification. In low-gain regimes characteristic of most experimental realizations, including ours, the probability distribution is approximately independent of the gain (Eq. 1 and supplementary text, section B). This allowed us to map the ground-state

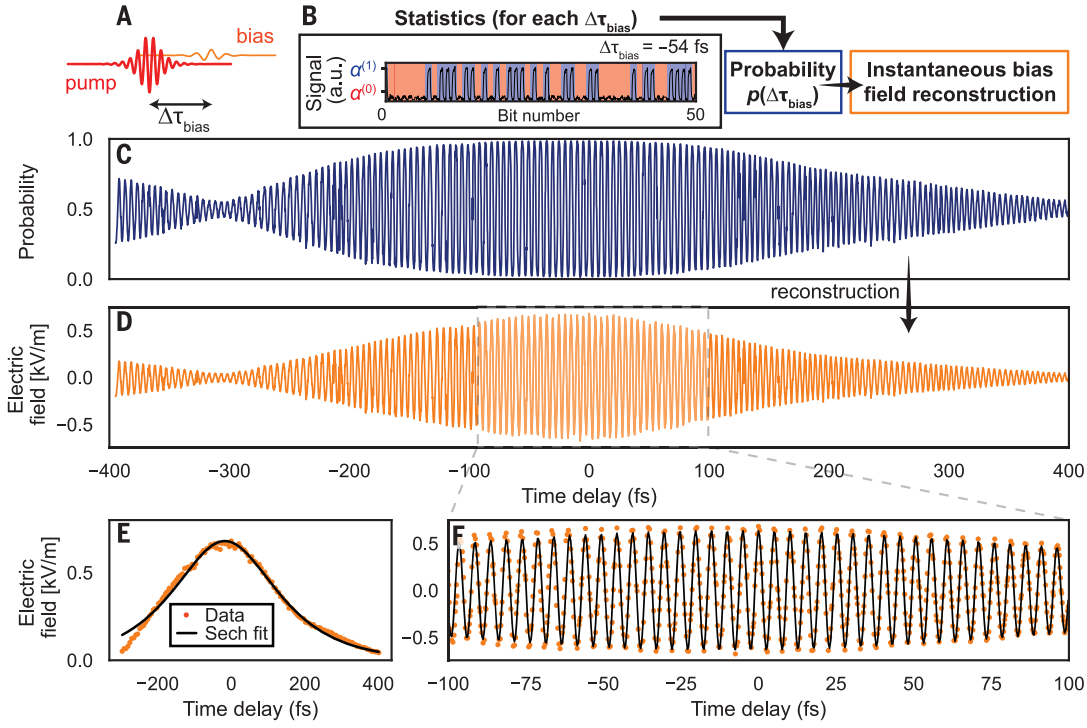


Fig. 4. Measuring sub–photon-level ultrashort pulses. (**A**) Schematic of the bias time-delay control in this experiment. (**B**) Field reconstruction method. (**C**) Measured probability as a function of the time delay $\Delta\tau_{bias}$. (**D**) Corresponding reconstructed bias field. (**E**) Extracted field envelope and corresponding sech fit. (**F**) Central part of the pulse and corresponding fit.

wave function density $|\Psi_0(E_{vac})|^2$ to the bias sensitivity of the probability dp/db (Fig. 3D). The reconstructed $|\Psi_0(E_{vac})|^2$ closely matches the expected distribution from quantum optics $|\Psi_0(E_{vac})|^2 \propto \exp\left[-E^2/\left(2\Delta E_{vac}^2\right)\right]$. This result is another strong indication that biased vacuum fluctuations are at the origin of the randomness in our experiment.

The probability's sensitivity to the phase bias can be directly used to sense the temporal dependence of sub–photon-level fields (fields for which the mean number of photons in the pulse is less than one). In the schematic of our field-sensing experiment (Fig. 4A), the bias field was time-delayed with respect to the pump. At each $\Delta\tau_{bias}$, we recorded statistics of the OPO phase (Fig. 4B). By inverting the bias-probability relationship, we could estimate the time-dependent bias field: $b(\Delta\tau_{bias}) \propto \mathrm{erf}^{-1}[2p(\Delta\tau_{bias}) - 1]$, where erf^{-1} is the inverse of the erf function.

The reconstructed field $b(\Delta\tau_{bias})$ (Fig. 4D) reproduces the main characteristics of the bias' temporal line shape. From the field line shape, we could extract the pulse width and center wavelength (data and fits are shown in Fig. 4, E and F). Our analysis yielded estimated values for the main lobe of ~220 ± 6 fs and ~1520 ± 0.2 nm, respectively, which is in agreement with nominal values of the laser we used for this experiment. Specifically, the bias envelope is ~30 fs broader than the theoretical value, which can be attributed to a ~100-fs OPO signal at early times, which is consistent with the pump temporal width. We show the influence of OPO signal bandwidth and center frequency on our reconstruction method in the supplementary text, section C, and we discuss further justification of this reconstruction method in section B. Although we demonstrated bias and vacuum field reconstruction in this work, our method could in principle be used to sense the OPO signal by measuring the influence of a known bias field on the phase statistics. We expect this method to prove particularly useful in measuring transient quantum states of light that can be generated in OPO cavities (*39*).

Discussion and conclusions

Taken together, the methods proposed in this work suggest the use of vacuum-level bias fields as a probe of transient dynamics in nonlinear driven-dissipative systems with potential applications in sensing and computing. Our sensing method could be extended to observe the field dynamics in the cavity (Fig. 3D) by appropriately sweeping over the bias time delay and amplitude. Conversely, quantum states of light for biasing offers another degree of freedom to shape the bias-probability relationship.

An anticipated extension of our work will be to incorporate this photonic p-bit into a computing platform to realize photonic probabilistic computing (*27*). This can be realized by leveraging the extensive body of work on photonic Ising machines (*40*, *41*). The main speed limitations of our p-bit come from the control electronics and cavity mode lifetime (*20*). To increase the repetition rate of our photonic p-bit, one could resort to integrated photonics (*21*) or multimode laser cavities (*23*). Realizing our approach in multistable systems [for example, in integrated arrays of nanolasers (*42*)] paves the way to p-bits with extreme bandwidth, emulation of complex many-body Hamiltonians (by engineering couplings between p-bits), and nontrivial p-bit topologies (by using Hamiltonians H_0 with higher-order symmetries).

REFERENCES AND NOTES

1. R. Loudon, *The Quantum Theory of Light* (Oxford Univ. Press, 2000).
2. E. Purcell, *Phys. Rev.* **69**, 681 (1946).
3. W. E. Lamb Jr., R. C. Retherford, *Phys. Rev.* **72**, 241–243 (1947).
4. H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop, F. Capasso, *Science* **291**, 1941–1944 (2001).
5. V. Sandoghdar, C. I. Sukenik, E. A. Hinds, S. Haroche, *Phys. Rev. Lett.* **68**, 3432–3435 (1992).
6. C. M. Wilson *et al.*, *Nature* **479**, 376–379 (2011).
7. S. W. Hawking, *Nature* **248**, 30–31 (1974).
8. W. G. Unruh, *Phys. Rev. D* **14**, 870–892 (1976).
9. D. T. Smithey, M. Beck, M. G. Raymer, A. Faridani, *Phys. Rev. Lett.* **70**, 1244–1247 (1993).
10. C. Riek *et al.*, *Science* **350**, 420–423 (2015).
11. I.-C. Benea-Chelmus, F. F. Settembrini, G. Scalari, J. Faist, *Nature* **568**, 202–206 (2019).
12. A. Fragner *et al.*, *Science* **322**, 1357–1360 (2008).
13. I.-C. Hoi *et al.*, *Nat. Phys.* **11**, 1045–1049 (2015).
14. M. Kardar, *Statistical Physics of Fields* (Cambridge Univ. Press, 2007).
15. M. Herrero-Collantes, J. C. Garcia-Escartin, *Rev. Mod. Phys.* **89**, 015004 (2017).
16. P. D. Drummond, K. J. McNeil, D. F. Walls, *Int. J. Opt.* **27**, 321–335 (1980).
17. P. D. Drummond, K. J. McNeil, D. F. Walls, *Int. J. Opt.* **28**, 211–225 (1981).
18. C. Gabriel *et al.*, *Nat. Photonics* **4**, 711–715 (2010).
19. M. Jofre *et al.*, *Opt. Express* **19**, 20665–20672 (2011).
20. A. Marandi, N. C. Leindecker, K. L. Vodopyanov, R. L. Byer, *Opt. Express* **20**, 19322–19330 (2012).
21. Y. Okawachi *et al.*, *Opt. Lett.* **41**, 4194–4197 (2016).

22. T. Steinle, J. N. Greiner, J. Wrachtrup, H. Giessen, I. Gerhardt, *Phys. Rev. X* **7**, 041050 (2017).
23. K. Kim *et al*., *Science* **371**, 948–952 (2021).
24. M. Gendreau, J.-Y. Potvin, Eds. *Handbook of Metaheuristics* (Springer, ed. 2, 2010).
25. C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach, D. Blei, Eds. (PMLR, 2015), pp. 1613–1622.
26. Z. Ghahramani, *Nature* **521**, 452–459 (2015).
27. K. Y. Camsari, B. M. Sutton, S. Datta, *Appl. Phys. Rev.* **6**, 011305 (2019).
28. D. Marković, A. Mizrahi, D. Querlioz, J. Grollier, *Nat. Rev. Phys.* **2**, 499–510 (2020).
29. A. Nichol *et al*., GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models. arXiv:2112.10741 [cs.CV] (2021).
30. W. A. Borders *et al*., *Nature* **573**, 390–393 (2019).
31. N. A. Aadit *et al*., *Nat. Electron.* **5**, 460–468 (2022).
32. A. Z. Pervaiz, B. M. Sutton, L. A. Ghantasala, K. Y. Camsari, *IEEE Trans. Neural Netw. Learn. Syst.* **30**, 1920–1926 (2019).
33. K. E. Harabi *et al*., *Nat. Electron.* **6**, 52–63 (2023).
34. G. Wetzstein *et al*., *Nature* **588**, 39–47 (2020).
35. N. P. Barnes, K. E. Murray, G. H. Watson, in *Advanced Solid State Lasers*, L. Chase, A. Pinto, Eds. vol. 13 of *OSA Proceedings Series* (Optica Publishing Group, 1992), paper NC6.
36. H. Stover, W. Steier, *Appl. Phys. Lett.* **8**, 91–93 (1966).
37. C. J. Buczek, R. J. Freiberg, M. L. Skolnick, *Proc. IEEE* **61**, 1411–1431 (1973).
38. Materials and methods are available as supplementary materials.
39. R. Y. Teh *et al*., *Phys. Rev. A* **101**, 043807 (2020).
40. A. Marandi, Z. Wang, K. Takata, R. L. Byer, Y. Yamamoto, *Nat. Photonics* **8**, 937–942 (2014).
41. N. Mohseni, P. L. McMahon, T. Byrnes, *Nat. Rev. Phys.* **4**, 363–379 (2022).
42. M. Parto, W. Hayenga, A. Marandi, D. N. Christodoulides, M. Khajavikhan, *Nat. Mater.* **19**, 725–731 (2020).
43. C. Roques-Carmes *et al*., charlesrc/pbit-opo: First release. Zenodo (2023); https://doi.org/10.5281/zenodo.7972106.

ACKNOWLEDGMENTS

The authors thank A. Heiner, J. Mastron, and A. Wiesboeck from Toptica for their help in setting up the dual-wavelength femtosecond laser. The authors also acknowledge administrative support from J. Freedman. The authors acknowledge stimulating conversations with S. Z. Uddin, C. Spägele, R. Dangovski, D. Luo, S. Kim, Z. Lin, T. Pezeril, and E. Ippen. **Funding:** Y.S. acknowledges support from the Swiss National Science Foundation (SNSF) through the Early Postdoc Mobility Fellowship P2EZP2-188091. J.S. acknowledges support from a Mathworks fellowship. S.C. acknowledges support from a Korea Foundation for Advanced Studies Overseas PhD Scholarship. Research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-18-2-0048. **Authors contributions:** C.R.-C., Y.S., and M.S. conceived the original idea. C.R.-C. and Y.S. built the experimental setup, with contributions from S.C., G.V., E.K., and S.E.K.; C.R.-C., Y.S., J.S., and S.C. acquired and analyzed the data. J.S. developed the theoretical and numerical tools, with contributions from C.R.-C. and N.R.; J.D.J. and M.S. supervised the project. The manuscript was written by C.R.-C., Y.S., and J.S., with inputs from all authors. **Competing interests:** The authors declare no potential competing financial interests. **Data and materials availability:** All data are available in the manuscript, the supplementary materials, or deposited online (*43*). Correspondence and requests for materials should be addressed to C.R.-C. (chrc@stanford.edu) and Y.S. (salamin@mit.edu).
License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adh4920
Materials and Methods
Supplementary Text
Figs. S1 to S12
Table S1
References (*44*–*59*)

Submitted 7 March 2023; accepted 6 June 2023
10.1126/science.adh4920

SOLAR CELLS

Engineering ligand reactivity enables high-temperature operation of stable perovskite solar cells

So Min Park[1]†, Mingyang Wei[2]†, Jian Xu[1]†, Harindi R. Atapattu[3], Felix T. Eickemeyer[2], Kasra Darabi[4], Luke Grater[1], Yi Yang[5], Cheng Liu[5], Sam Teale[1], Bin Chen[1,5], Hao Chen[1], Tonghui Wang[4], Lewei Zeng[1], Aidan Maxwell[1], Zaiwei Wang[1], Keerthan R. Rao[3], Zhuoyun Cai[6], Shaik M. Zakeeruddin[2], Jonathan T. Pham[6], Chad M. Risko[3], Aram Amassian[4], Mercouri G. Kanatzidis[5], Kenneth R. Graham[3]*, Michael Grätzel[2]*, Edward H. Sargent[1,5,7]*

Perovskite solar cells (PSCs) consisting of interfacial two- and three-dimensional heterostructures that incorporate ammonium ligand intercalation have enabled rapid progress toward the goal of uniting performance with stability. However, as the field continues to seek ever-higher durability, additional tools that avoid progressive ligand intercalation are needed to minimize degradation at high temperatures. We used ammonium ligands that are nonreactive with the bulk of perovskites and investigated a library that varies ligand molecular structure systematically. We found that fluorinated aniliniums offer interfacial passivation and simultaneously minimize reactivity with perovskites. Using this approach, we report a certified quasi–steady-state power-conversion efficiency of 24.09% for inverted-structure PSCs. In an encapsulated device operating at 85°C and 50% relative humidity, we document a 1560-hour T_{85} at maximum power point under 1-sun illumination.

Metal-halide perovskite solar cells (PSCs) are emerging photovoltaic (PV) technologies that hold promise in terawatt-scale deployment (*1*). They unite high power-conversion efficiencies (certified PCEs up to 25.7%) (*2*) with low-cost solution processing that uses abundant materials (*3*–*5*). To compete with crystalline silicon (c-Si) solar cells and to be applied with c-Si in tandem cells, PSCs will need improved evidence of bankability (*6*), such as operating stability at elevated temperatures in accelerated aging (*7*–*12*).

Interfaces, which have high defect densities (*13*, *14*) and low barriers to ion transfer (*15*, *16*) and are susceptible to moisture- and oxygen-induced degradation (*17*, *18*), can provide pathways for energy loss and degradation in PSCs (*19*, *20*). In record-efficiency PSCs, interfaces have been passivated by using two- and three-dimensional (2D/3D) hybrid structures (*21*–*23*), where a thin layer of Ruddlesden-Popper 2D perovskites terminates the 3D-perovskite surface. These 2D/3D structures are constructed by exposing perovskite surfaces to a solution containing ammonium ligands, during which the 3D lattice is fragmented into 2D layers (*24*, *25*). This approach is beneficial because the 2D overlayer increases the resistance of perovskites to degradation (*17*, *26*).

There is a diversity of findings in reports on the stability of 2D overlayers under stress. Some studies have indicated limited stability of the 2D/3D interface under thermal stress (*27*–*29*). Progress has been made, with a recent study demonstrating 500-hour operating stability under conditions of 65°C, 50% relative humidity (RH), and maximum power point (MPP) tracking [International Summit on Organic Photovoltaic Stability (ISOS)-L-3-65°C], enhanced by the incorporation of 3-fluorophenethylammonium (3FPEA) intercalation (*23*). Another study (*30*) that used an oleylammonium-intercalated 2D overlayer showed promising results in a 1000-hour damp-heat test, although this study was based on silicon PV module IEC 61215:2016 standards (85°C and 85% RH, dark).

Given these promising performance and stability improvements, the field continues to pursue ever-higher durability targets. There is interest in raising the standard to 85°C ISOS-L-3 (MPP tracking at 85°C and 50% RH). However, the reactivity of ammonium ligands with 3D perovskites may lead to further penetration into the bulk perovskite film (*28*, *31*, *32*) and may contribute to deterioration in device performance under these very demanding stress conditions (*27*, *28*, *31*–*33*). Thus, we explored noninvasive surface-passivating ligands, but we did note that in prior studies, these have yet to achieve the remarkable benefits of the 2D/3D strategy.

We sought to understand how molecular structure affects ligand reactivity with 3D perovskites. Previous studies used bul[illegible] ammonium ligands to suppress ligand in[illegible]calation, whereas certain small-sized ammonium ligands exhibit low reactivity (*34*, *35*). We attribute this difference to the limited availability of characterization techniques that allow one to compare ligand reactivity among different ammonium molecules—a particularly challenging pursuit in view of the ultrathin nature (typically <10 nm) of the passivating overlayer (*23*).

Spectroscopy studies of ammonium ligands in perovskites

We used angle-resolved x-ray photoelectron spectroscopy (AR-XPS) and time-of-flight secondary ion mass spectrometry (TOF-SIMS) to investigate the penetration of a library of ammonium ligands into the bulk of perovskites (Fig. 1A), which include varying tail groups and alkyl chain lengths. A small-sized ammonium ligand, anilinium (An), had the lowest reactivity with 3D perovskites. We then explored derivatives of An with different degrees of fluorination to couple low ligand reactivity with effectual interfacial passivation. These molecules improved operating stability for the encapsulated PSC at 85°C and 50% RH.

Ammonium ligands were deposited on $Cs_{0.05}MA_{0.05}FA_{0.9}Pb(I_{0.95}Br_{0.05})_3$ perovskite films (materials and methods), and films were annealed at 100°C (*31*). We used AR-XPS to probe the spatial distribution of ammonium ligands in the out-of-plane direction (Fig. 1B). AR-XPS is a nondestructive depth-profiling technique used to determine the composition of ultrathin layers (<10 nm) at the top surfaces of films (*36*, *37*). We varied the probe depth by changing the angle between the normal of the perovskite sample and the analyzer. Three different electron takeoff angles (0°, 45°, and 75°) were selected, with probe depths varying from ~6 to 8 nm at 0° to ~1 to 2 nm at 75° (Fig. 1C).

To obtain the proportion of ammonium ligands at a given probe depth, we compared the ratio of C-N, where N signals originated from methylammonium (MA) and ammonium ligands at a binding energy of ~401.5 eV, with the N signals originating from formamidinium (FA) at a binding energy of ~399.8 eV in the peak area in the N 1s XPS spectrum (figs. S1 and S2). The C-N/FA N ratio plots (Fig. 1D) revealed that the aryl ammoniums phenethylammonium (PEA) and 3FPEA had relatively uniform depth distributions. By contrast, alkyl ammoniums butylammonium (BA) and octylammonium (OA) tended to accumulate on the top surfaces, as evidenced by the C-N/FA N ratio increasing with larger electron takeoff angles. However, each of the ligands yielded C-N/FA N ratios greater than those of untreated perovskites (control) for all electron

[1]Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario M5S 3G4, Canada. [2]Laboratory of Photonics and Interfaces, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. [3]Department of Chemistry, University of Kentucky, Lexington, KY 40506, USA. [4]Department of Materials Science and Engineering, and Organic and Carbon Electronics Laboratories (ORaCEL), North Carolina State University, Raleigh, NC 27695, USA. [5]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [6]Department of Chemical and Materials Engineering, University of Kentucky, Lexington, KY 40506, USA. [7]Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208, USA.
†These authors contributed equally to this work.
*Corresponding author. Email: ted.sargent@utoronto.ca (E.H.S.); michael.graetzel@epfl.ch (M.G.); kenneth.graham@uky.edu (K.R.G.)

takeoff angles (Fig. 1E), indicating their distinct contributions to C-N signals and their intercalation into the bulk of perovskites.

We then examined long-chain alkyl ammonium decylammonium (DA), along with small-sized An (*34*, *35*, *38*). For DA, the C-N/FA N ratio decreased with a larger electron takeoff angle (Fig. 1D). The reversed trend, as compared with other alkyl ammoniums, suggested that DA molecules diffused into the bulk of perovskites. By contrast, An exhibited low reactivity and had limited penetration into perovskites. The C-N/FA N ratios for An were a factor of ~2 less than those for PEA at different electron takeoff angles and are similar to control perovskites (Fig. 1E).

At these low C-N/FA N ratios, quantifying the extent of ligand penetration was difficult, given the prominent C-N contribution from MA cations (fig. S2 and table S1). We thereby performed TOF-SIMS and found that, whereas PEA cations were distributed throughout the thickness of the perovskite film, An cations were detected only at the top surface of the film (fig. S3). The combination of AR-XPS and TOF-SIMS supported An remaining surface localized.

Although An exhibited low ligand reactivity, it has also been reported to have limited passivating capacity (*35*, *39*). To promote interfacial passivation, we explored fluorinated An, including 4-fluoroanilinium (4FAn), 2,6-difluoroanilinium (26FAn), and 3,4,5-trifluoroanilinium (345FAn), as well as tert-butyl-substituted 3,5-di-tert-butylanilinium (35tbuAn) (*17*, *40*, *41*). The C-N/FA N ratios (Fig. 1E) revealed that, compared with An, the newly synthesized ligands retained similarly low reactivity with perovskites. For fluorinated An, this low reactivity was also seen in the C-F signal in the F 1s XPS spectrum (fig. S4). We identified a distinct C-F peak for 3FPEA, whereas the C-F signals from 4FAn, 26FAn, and 345FAn were all below the detection limit at the different electron takeoff angles. TOF-SIMS further confirmed that 345FAn cations were localized to the surface of the perovskite film (fig. S5).

We then characterized the phase transformation of perovskites driven by their interaction with ammonium ligands. Ultrafast transient reflection (TR) spectroscopy was used initially to detect 2D perovskite formation (*42*). For the PEA-treated perovskite films, negative reflectance features associated with n = 1, 2, and 3 layered 2D perovskites were observed in the TR spectra (Fig. 2A). The relative proportion of 2D phases increased as the solution exposure time was prolonged (fig. S6) and was indicative of progressive phase transformation. For the An- and 345FAn-treated perovskite films, no 2D phases were observed (Fig. 2, B and C), which we attributed to their inability to convert 3D perovskites into 2D phases (fig. S7). These results reinforce the model proposing that inhibiting 2D phase conversion would also reduce ligand penetration.

We also investigated phase transformation with grazing incidence wide-angle x-ray scattering (GIWAXS) measurements. Consistent with the TR studies, n = 2 2D perovskites were observed at around $q = 0.29$ Å^{-1} along q_z for the PEA-treated perovskite film (Fig. 2D) that blocks vertical carrier transport (*32*). Incidence angle–dependent diffraction patterns of PEA-treated perovskites revealed that the 2D phase formation was more pronounced near the surface (fig. S8). No 2D phase was detected in the An- and 345FAn-treated perovskites (Fig. 2, E and F).

In addition, we observed the formation of PbI_2 at $q = 0.9$ Å^{-1} for An-treated perovskite films,

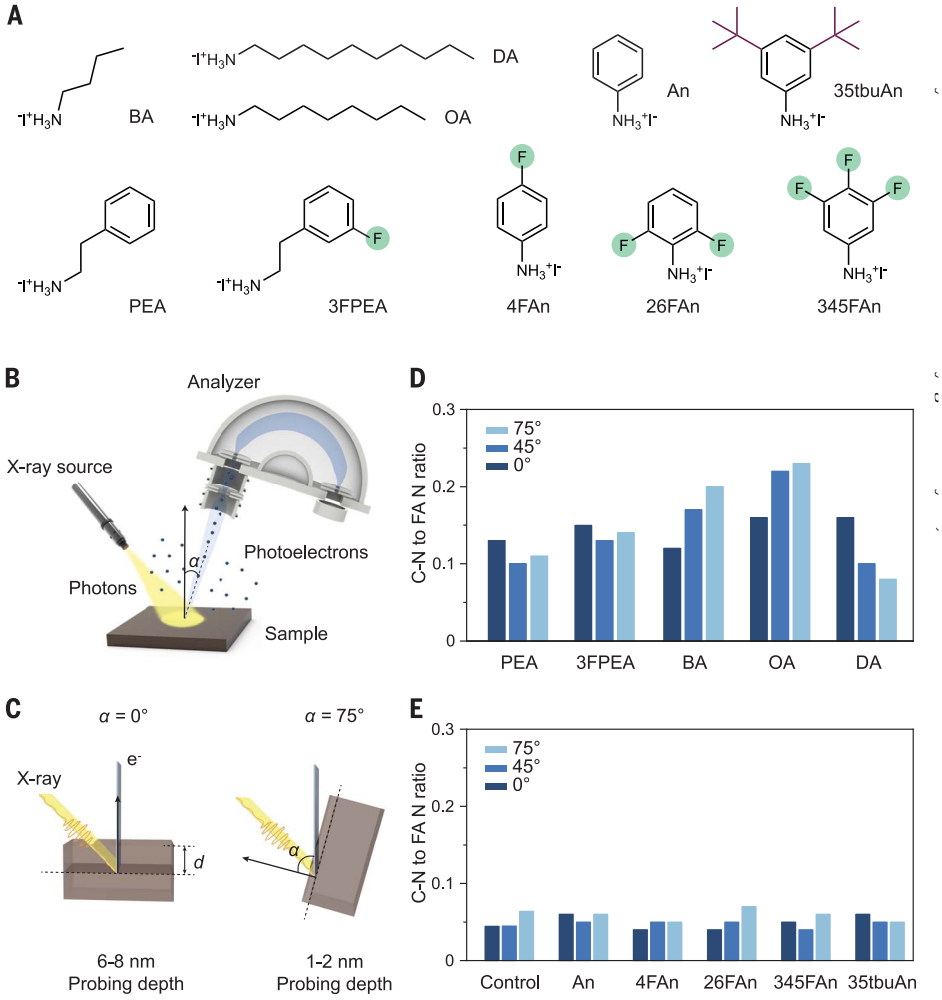


Fig. 1. AR-XPS characterization of ammonium ligand penetration. (**A**) Chemical structures of the ammonium ligands investigated in this study. (**B**) Experimental setup for AR-XPS characterization. The electron takeoff angle α is defined as the angle between the normal of the sample substrate and the analyzer. (**C**) Schematic indicating the relation between α and the photoelectron probing depth, *d*. (**D** and **E**) C-N to FA N ratios at electron takeoff angles of 0°, 45°, and 75° for the control (untreated) and ammonium ligand–treated perovskite films. The results show reduced bulk and surface existence for An and An derivatives.

which we attributed to polar solvent–induced decomposition during solution exposure, as seen in x-ray diffraction (XRD) characterization (fig. S9). This is a topic that requires further in-depth investigation. GIWAXS characterization showed that 345FAn protected the underlying perovskites from this solvent-induced degradation (Fig. 2F). We associate the reduction in surface degradation with increased hydrophobicity of this fluorinated ligand (fig. S10) (*26*).

Theoretical calculations

The interactions of ammonium ligands and perovskites are depicted in Fig. 3A. We performed density functional theory (DFT) calculations to explore how molecular structure affected these interactions. We first calculated the binding energies (E_b) of two adjacent perovskite fragments at their interface with the insertion of ammonium ligands (Fig. 3B). A more negative E_b implies that it is thermodynamically more favorable to form the 2D/3D heterostructures (*24*). Consistent with the AR-XPS studies, penetrating PEA had a more negative E_b value than alkyl ammoniums OA and BA, whereas An and its derivatives showed the least tendency to form 2D/3D structures (Fig. 3C).

It has been suggested that the formation of 2D perovskites is related to the steric hindrance around the NH_3^+ group (*43*). We calculated the steric effect index (STEI, defined as the steric environment around the NH_3^+ group) of the ammonium ligand (*43*) and found that An had a much larger STEI (2.34) than that of PEA (1.21). Furthermore, phase transformation entailed the replacement of A-site cations at 2D/3D interfaces (*24*). DFT calculations indicated that this process demanded more energy for An-intercalated interfaces than for PEA-intercalated ones (fig. S11). This evidence further supported our finding that An was less likely to penetrate the bulk of perovskites through ligand intercalation.

We proceeded to examine the interactions between ammonium ligands and perovskite surfaces. We calculated the interaction energy (E_{int}) for the scenario in which an ammonium ligand occupied an MA-vacancy site on the perovskite surface (Fig. 3D). Electrostatic potential calculations showed that fluorine substitution resulted in a higher positive charge density near the ammonium group (fig. S12) that could enhance its binding with the negatively charged MA vacancy (*13*). Indeed, the E_{int} values were –0.74, –0.89, –0.92, and –1.22 eV for An, 4FAn, 26FAn, 345FAn, respectively (Fig. 3E), indicating that 345FAn exhibited the strongest interaction with the perovskite surface. Future computational studies could beneficially predict the binding energy of the N 1s XPS peak.

Perovskite degradation and solar cell studies

We used TR and TOF-SIMS to study perovskite degradation. The TR results indicated that PEA-based 2D perovskites were thermally unstable on the film surface and decomposed into PbI_2 after thermal aging at 85°C for 2 hours (Fig. 4, A and B). This process we associated with the diffusion of PEA cations into bulk perovskites,

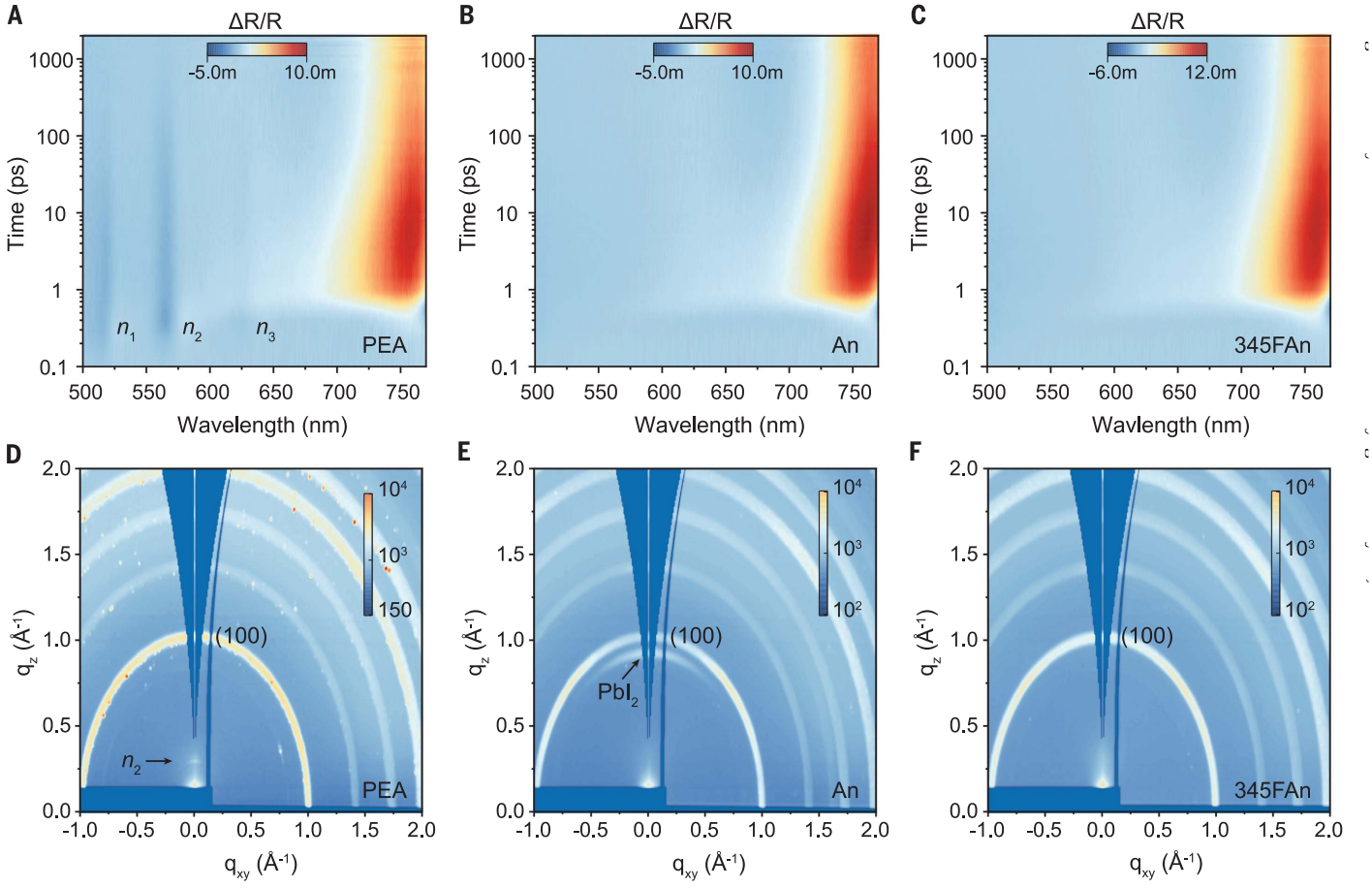


Fig. 2. Phase transformation of perovskite films. (**A** to **C**) Pseudocolor representation of the transient reflectance spectra, $\Delta R/R$, for the PEA- (A), An- (B), and 345Fan-treated (C) perovskite films, respectively. n = 1, n = 2, n = 3 2D perovskites (n_1, n_2, and n_3) appear in the film after PEA passivation. (**D** to **F**) GIWAXS image for perovskite films of PEA (D), An (E), and 345FAn (F) passivation, respectively. The color bar shows the diffraction intensity collected from the GIWAXS detector. q_{xy} and q_z represent in-plane and near out-of-plane scattering vectors, respectively. n = 2 2D perovskite (n_2) and PbI_2-related diffraction peaks are observed in the PEA- and An-treated films, respectively. (100) refers to the (100) diffraction peak of 3D perovskites.

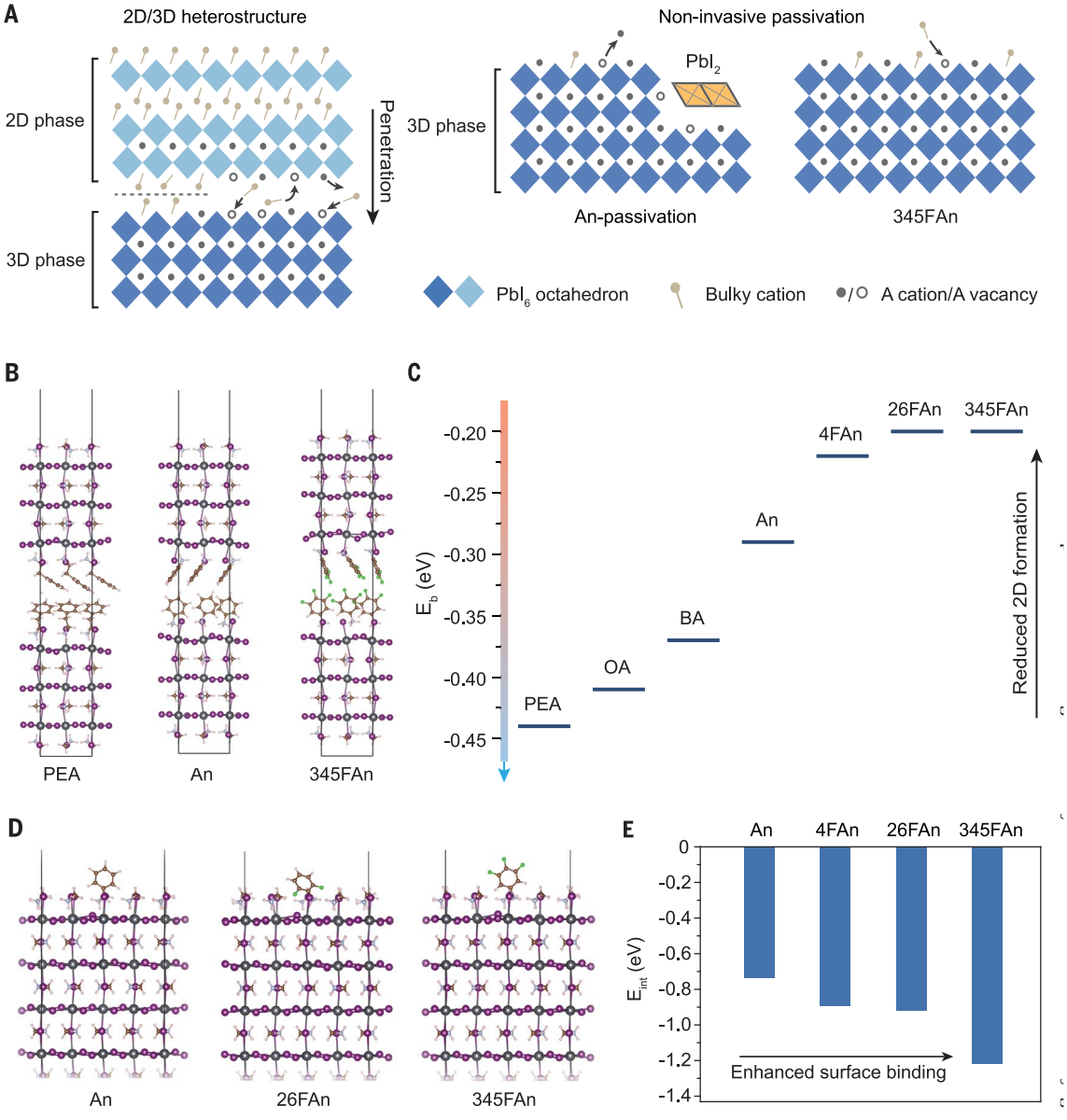


Fig. 3. DFT studies. (**A**) Schematic depiction of the model of ammonium ligand and perovskite interactions. (**B**) Models used in DFT calculations to calculate the binding energy (E_b) of adjacent 2D perovskite fragments. (**C**) E_b values for seven ammonium ligands substituted into $MAPbI_3$ perovskite slabs. (**D**) Models used to calculate the interaction energy (E_{int}) of ammonium ligands with the $MAPbI_3$ perovskite surface. (**E**) E_{int} values of four ammonium ligands with the perovskite surface.

as was seen with TOF-SIMS (Fig. 4C) and AR-XPS (fig. S13). The dynamic nature of 2D/3D heterostructures was also confirmed for BA, OA, DA, and 3FPEA (fig. S14). By contrast, neither ligand penetration nor phase degradation was observed for An- and 345FAn-treated films (Fig. 4C and fig. S15). These results suggested that suppressing ligand intercalation into 3D perovskites improved interface stability under thermal stress.

We monitored the photoluminescence (PL) stability of perovskite films after annealing at 85°C (Fig. 4D). The emission intensity of the untreated control and PEA-treated films degraded to 30% of its initial value within 144 hours (Fig. 4E), as seen in the previous reports (*31*). By contrast, An- and 345FAn-treated films show improved thermal stability, and 345FAn-treated perovskites retained 85% of their initial brightness after annealing.

We fabricated inverted-structure PSCs to investigate how ammonium ligands influenced the efficiency and stability of PSCs. We used both $Cs_{0.05}MA_{0.05}FA_{0.9}Pb(I_{0.95}Br_{0.05})_3$ and $Cs_{0.05}MA_{0.15}FA_{0.8}PbI_3$ perovskites as the absorber, the self-assembled monolayer 2PACz as the hole transport layer, the thermally evaporated C_{60}/bathocuproine (BCP) bilayer as the electron transport layer, and indium tin oxide (ITO) as the transparent electrode (materials and methods). Three representative ammonium ligands, including the 2D-forming PEA cation and the noninvasive An and 345FAn cations, were used for interfacial engineering. The cross-sectional scanning electron microscope (SEM) image of a complete device is shown in fig. S16.

Device performance (Fig. 5A) showed that compared with control devices, PEA-treated $Cs_{0.05}MA_{0.15}FA_{0.8}PbI_3$ PSCs (16 devices) provided improved PV performance with an average PCE of 23.2% (Fig. 5A and fig. S17). The average PCE of An devices decreased to 19.9%, a finding we linked to perovskite decomposition. Because of the improved surface passivation, 345FAn treatment increased average PCEs of $Cs_{0.05}MA_{0.15}FA_{0.8}PbI_3$ and $Cs_{0.05}MA_{0.05}FA_{0.9}Pb(I_{0.95}Br_{0.05})_3$ PSCs to 23.3 and 22.9%, respectively (Fig. 5A and fig. S18). The external quantum efficiency (EQE) spectra of the champion cells (Fig. 5B) for PEA and 345FAn devices exhibited improved charge collection compared with controls, and their integrated short-circuit current density (J_{sc}) values matched well with those from the current-voltage (I-V) sweep.

We sent a 345FAn-treated $Cs_{0.05}MA_{0.15}FA_{0.8}PbI_3$ device to National Renewable Energy Laboratory (NREL, USA) for independent characterization. The device delivered a certified PCE of 24.09% by using a quasi–steady-state (QSS) I-V

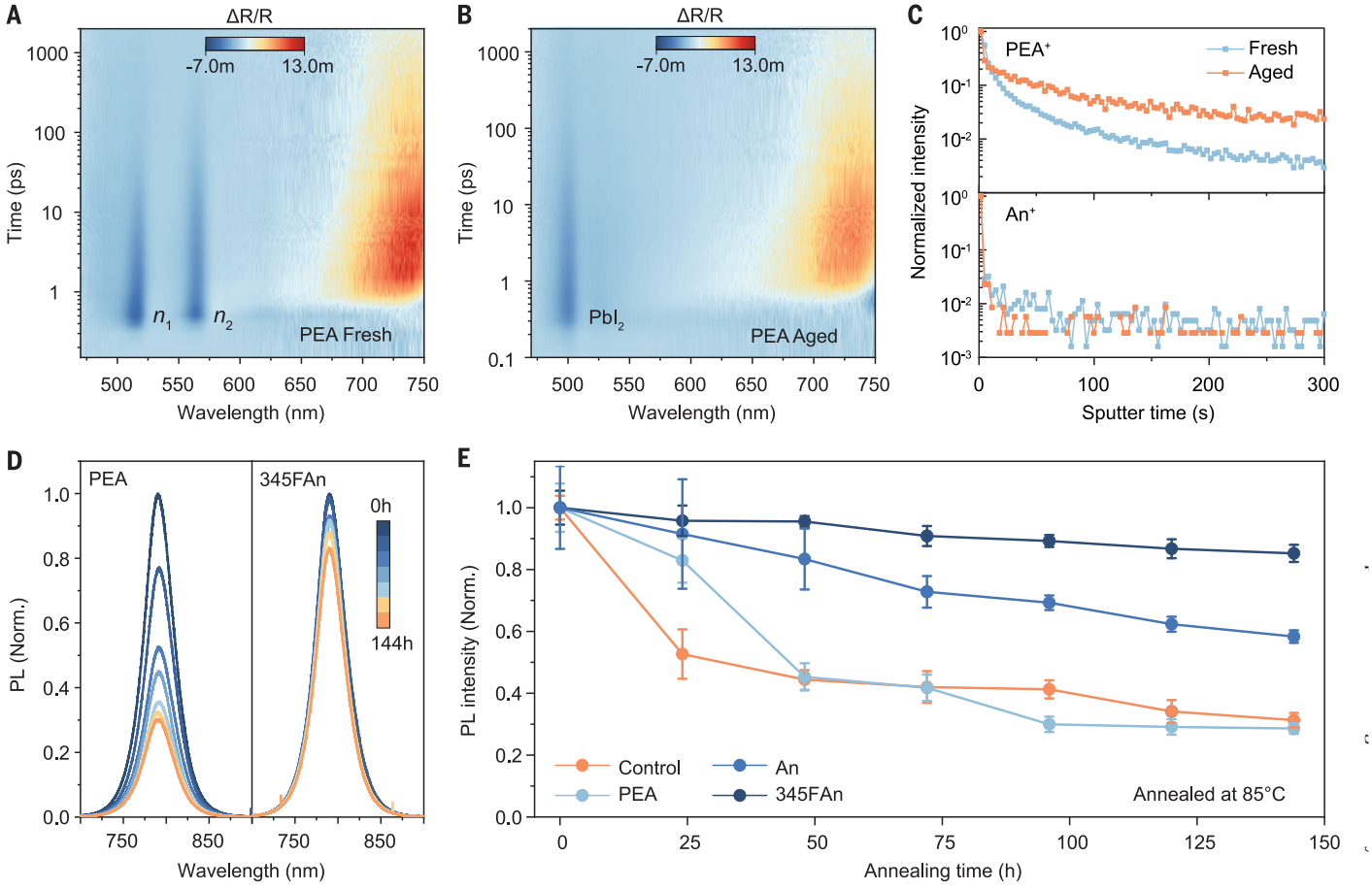


Fig. 4. Thermal stability of perovskite films. (**A** and **B**) Pseudocolor representation of the transient reflectance spectra for PEA-treated perovskite films before (A) and after (B) thermal aging at 85°C for 2 hours. 2D perovskites were decomposed into PbI_2 after thermal aging. (**C**) The distribution of PEA^+ and An^+ in fresh and aged (at 85°C for 2 hours) perovskite films observed with TOF-SIMS. (**D**) PL spectra of perovskite thin films under thermal annealing at 85°C. The films were kept on a hot plate at 85°C in the N_2-filled glovebox and measured every 24 hours in ambient air. (**E**) Changes in PL intensity for different annealing times. Data are presented as mean values ± standard deviation. Three films were measured under each condition.

sweep (Fig. 5C and fig. S19). The QSS-certified efficiency reported for inverted PSCs is compared with that of other devices in table S2.

Seeking a better understanding of the improved performance, we used ultraviolet photoelectron spectroscopy (UPS) to characterize the electronic structure of the perovskite films. The secondary electron cutoff spectra (fig. S20) showed that 345FAn decreased the work function of perovskites from 4.64 to 4.17 eV and shifted the valence-band maximum from 1.15 eV below the Fermi level in the control perovskites to 1.50 eV below the Fermi level. These results show that 345FAn induced more n-type character in the perovskite film (fig. S20), which in inverted PSCs leads to favorable band bending for electron extraction and a decrease in nonradiative carrier recombination (*44*, *45*).

We performed optoelectronic characterization to investigate carrier recombination. From photoluminescence quantum yield (PLQY) measurements, we identified nonradiative recombination at the perovskite/C_{60} interface as the limiting factor for the performance of control devices (fig. S21). This recombination loss corresponded to a reduction of the quasi-Fermi level splitting (QFLS) by ~100 meV (table S3) (*46*, *47*). We found that PEA and 345FAn passivation could reduce the interfacial nonradiative recombination, as reflected by the improved QFLS relative to the control stack, by ~15 and ~30 meV, respectively. We further analyzed the energy loss of PSCs on the basis of the device diode characteristics (figs. S22 and S23) (*48*). Consistent with the PL studies, nonradiative losses were reduced for PEA and 345FAn devices (12.7 and 12.1% for PEA and 345FAn, respectively) compared with those of the control device (16.5%) (fig. S24).

We evaluated the effect of ligand reactivity on the material processing of PSCs. We varied the solution exposure time between 0 s (dynamic spinning) and 120 s and tracked the PV parameters of PSCs accordingly (Fig. 5D and fig. S25). The average PCE of PEA devices (eight devices for each condition) dropped to 15.8%, with an average fill factor (FF) of merely 67.4% over the course of solution exposure for 30 s, whereas 345FAn devices maintained an average PCE of 23.0% (Fig. 5D). Wide solution-processing windows were also achieved for other nonpenetrating ligands, such as 35tbuAn, 4FAn, and 26FAn (fig. S26). In pursuit of initial evidence of capacity to scale area, we fabricated perovskite solar modules (PSMs) with an active area of 22 cm^2 and nine interconnected subcells (materials and methods). The champion PCE under reverse voltage scan was improved from 19.9 to 20.8% for the 345FAn-treated PSMs, with an open-circuit voltage (V_{oc}) of 10.13 V, an FF of 74.2%, and a J_{sc} of 2.77 mA cm^{-2} (fig. S27).

We studied the operating stability of PSCs using ISOS protocols (*49*). Accelerated lifetime testing was performed at the ISOS-L-3-85°C level, constituting light-soaking tests at 50% RH and 85°C with MPP tracking. We used atomic-layer deposition for SnO_2 as the buffer layer and $Cs_{0.05}MA_{0.05}FA_{0.9}Pb(I_{0.95}Br_{0.05})_3$ perovskites as the absorber (materials and methods). The initial PCEs of the encapsulated control, PEA, and 345FAn devices were 18.9, 20.1, and 20.2%, respectively (fig. S28). The stability

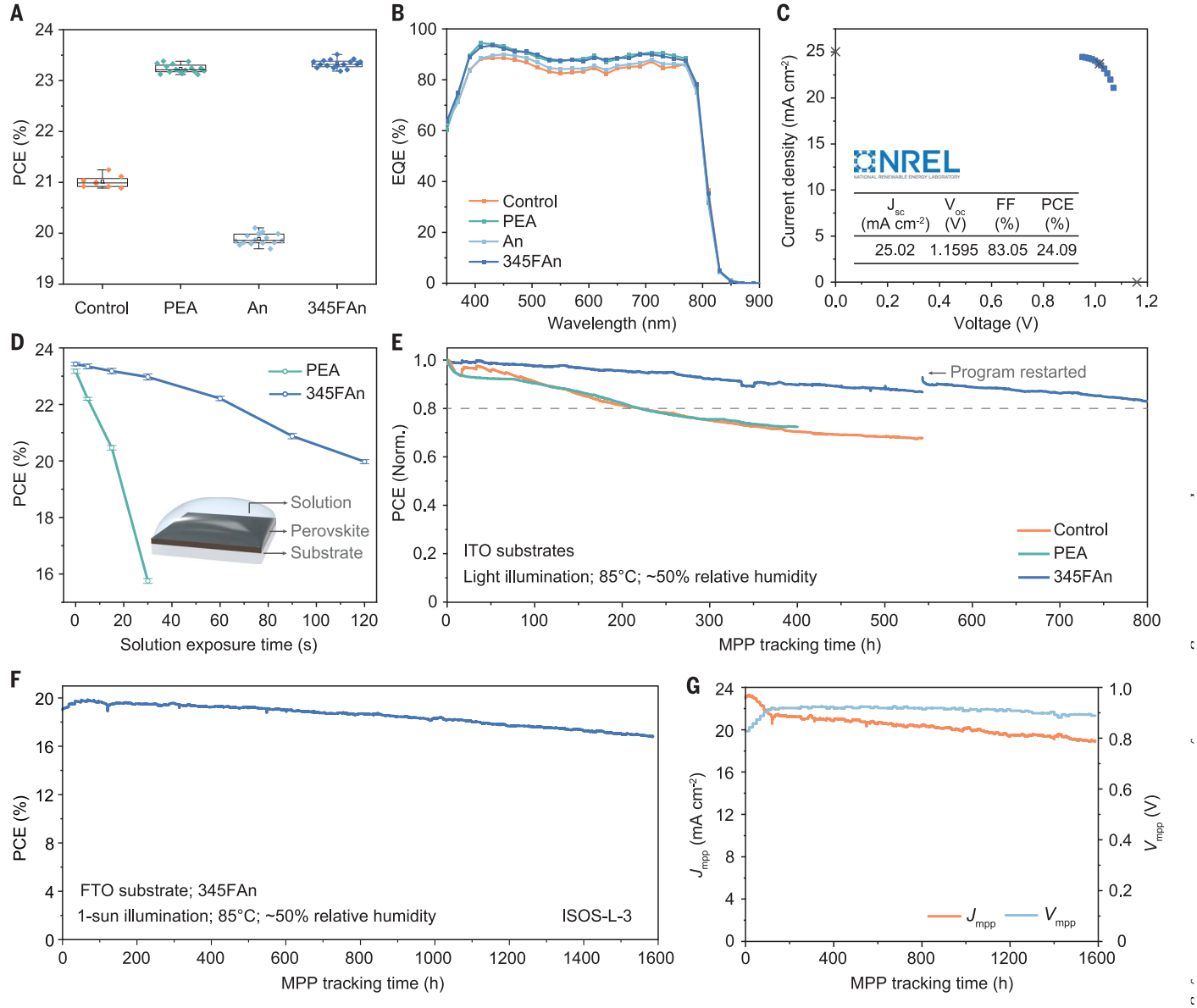


Fig. 5. Photovoltaic performance of PSCs with interface engineering. (**A**) PCE for control (8 devices) versus ammonium ligand–treated (16 devices for each type) PSCs. (**B**) EQE curves of the control and ammonium ligand–treated devices. (**C**) QSS current density-voltage (*J-V*) curve of one representative 345FAn device certified at NREL. (Inset) PV parameters of the device. (**D**) PCE evolution of PEA and 345FAn devices as a function of solution exposure time (8 devices for each condition). Data are presented as mean values ± standard deviation. (Inset) Schematic illustration of the solution exposure process for interface engineering. (**E**) MPP tracking of encapsulated control, PEA, and 345FAn devices performed using ITO substrates at 85°C with an RH of ~50% under 0.8-sun illumination. (**F** and **G**) PCE (F), current density (J_{mpp}), and voltage (V_{mpp}) (G) at the MPP of the encapsulated 345FAn device determined using an FTO substrate under 1-sun illumination at 85°C and ~50% RH.

results are shown in Fig. 5E. We recorded a similar T_{80} of ~220 hours for the control and PEA devices, in accordance with the limited thermal stability of perovskite films (Fig. 3E). We found that 2D/3D PSCs based on alkyl ammonium BA and OA were even less stable than the control device (fig. S29). By contrast, the 345FAn device retained >80% of its initial value after the 800-hour test. We estimate the T_{80} for the 345FAn device to be ~810 hours (fig. S30), representing a fourfold enhancement over the operating lifetime of the PEA device.

To improve the T_{80} further, we moved to fluorine-doped tin oxide (FTO) as an alternative transparent electrode (fig. S31) because of its chemical stability (*50*, *51*). We monitored the MPP of the encapsulated device under 1-sun illumination at 50% RH and 85°C. The initial PCE was 19.0%, which increased progressively to a peak value of 19.9% after ~50 hours of operation (Fig. 5F). This maximum PCE (PCE_{max}) corresponds to a temperature coefficient of −0.12% /°C relative to the room temperature device PCE (21.5%), which is consistent with reported results for inverted PSCs (*52*). After 1586 hours of continuous operation, the PCE dropped to 16.8% (84% of the PCE_{max}), primarily because of a reduction in current density (Fig. 5G). We determined the T_{85} to be ~1560 hours (Fig. 5F). A comparison with other ISOS-L-3-stable PSCs is provided in table S4.

Discussion

Two types of 2D/3D heterostructures have been reported for ISOS-L-3-stable PSCs: 3FPEA-intercalated 2D/3D perovskites (*23*) and

all-inorganic heterostructures (*12*). For all-inorganic PSCs, the T_{85} reached an impressive 4000 hours at 85°C, but with further room for progress in PCE (≤17% when measured at 85°C). 3FPEA-based PSCs have achieved PCEs exceeding 23% at room temperature, but stable operation over 500 hours was limited to 65°C.

Aided by spectroscopic techniques, including AR-XPS, TOF-SIMS, and TR, we see evidence here that certain 2D/3D heterostructures, such as those employing typical ammonium ligands including PEA, 3FPEA, BA, and OA, may evince thermal degradation, something we assign to the dynamic nature of the interfaces at elevated temperatures. Anilinium and its fluorinated derivatives offer a more robust interface structure correlated with limited penetration into the bulk of perovskites. Fluorinated An contributes interface passivation linked to strong interactions with perovskite surfaces. The resultant optimized 345FAn devices are among the most stable PSCs seen under ISOS-L-3 protocols at 85°C, achieving PCE of 19.9% during MPP tracking at this elevated operating temperature.

REFERENCES AND NOTES

1. N. M. Haegel *et al*., *Science* **364**, 836–838 (2019).
2. National Renewable Energy Laboratory, Best research-cell efficiency chart, revised 26 May 2023; https://www.nrel.gov/pv/cell-efficiency.html.
3. H. Min *et al*., *Nature* **598**, 444–450 (2021).
4. M. Kim *et al*., *Science* **375**, 302–306 (2022).
5. J. Jeong *et al*., *Nature* **592**, 381–385 (2021).
6. Z. Song *et al*., *Energy Environ. Sci.* **10**, 1297–1305 (2017).
7. Y. Zhao *et al*., *Nat. Energy* **7**, 144–152 (2022).
8. S. Bai *et al*., *Nature* **571**, 245–250 (2019).
9. Q. Cao *et al*., *Sci. Adv.* **7**, eabg0633 (2021).
10. Y. Liu *et al*., *Angew. Chem. Int. Ed.* **59**, 15688–15694 (2020).
11. Y.-H. Lin *et al*., *Science* **369**, 96–102 (2020).
12. X. Zhao *et al*., *Science* **377**, 307–310 (2022).
13. R. Lin *et al*., *Nature* **603**, 73–78 (2022).
14. X. Li *et al*., *Science* **375**, 434–437 (2022).
15. L. Chao *et al*., *Research* **2020**, 2616345 (2020).
16. H. Kim *et al*., *Nat. Commun.* **11**, 3378 (2020).
17. H. Zhu *et al*., *Adv. Mater.* **32**, 1907757 (2020).
18. S. Yang *et al*., *Science* **365**, 473–478 (2019).
19. J. Warby *et al*., *Adv. Energy Mater.* **12**, 2103567 (2022).
20. S. M. Park, A. Abtahi, A. M. Boehm, K. R. Graham, *ACS Energy Lett.* **5**, 799–806 (2020).
21. J. J. Yoo *et al*., *Energy Environ. Sci.* **12**, 2192–2199 (2019).
22. J. J. Yoo *et al*., *Nature* **590**, 587–593 (2021).
23. H. Chen *et al*., *Nat. Photonics* **16**, 352–358 (2022).
24. A. H. Proppe *et al*., *Nat. Commun.* **12**, 3472 (2021).
25. S. Sidhik *et al*., *Science* **377**, 1425–1430 (2022).
26. Y. Liu *et al*., *Sci. Adv.* **5**, eaaw2543 (2019).
27. J. Chakkamalayath, N. Hiott, P. V. Kamat, *ACS Energy Lett.* **8**, 169–171 (2023).
28. C. A. R. Perini *et al*., *Adv. Mater.* **34**, 2204726 (2022).
29. A. A. Sutanto *et al*., *Nano Lett.* **20**, 3992–3998 (2020).
30. R. Azmi *et al*., *Science* **376**, 73–77 (2022).
31. Q. Jiang *et al*., *Nat. Photonics* **13**, 460–466 (2019).
32. C. Liu *et al*., *Nat. Commun.* **12**, 6394 (2021).
33. S. Kumar, L. Houben, K. Rechav, D. Cahen, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2114740119 (2022).
34. M. M. Tavakoli *et al*., *Energy Environ. Sci.* **11**, 3310–3320 (2018).
35. S.-H. Lee *et al*., *ACS Energy Lett.* **6**, 1612–1621 (2021).
36. C. S. Fadley, *Prog. Surf. Sci.* **16**, 275–388 (1984).
37. B. Lv, T. Qian, H. Ding, *Nat. Rev. Phys.* **1**, 609–626 (2019).
38. Y. Wei *et al*., *Adv. Energy Mater.* **9**, 1900612 (2019).
39. D. Bi *et al*., *Nat. Commun.* **9**, 4482 (2018).
40. H. Zhu *et al*., *J. Am. Chem. Soc.* **143**, 3231–3237 (2021).
41. M. A. Hope *et al*., *J. Am. Chem. Soc.* **143**, 1529–1538 (2021).
42. Y. Yang *et al*., *Nat. Energy* **2**, 16207 (2017).
43. R. Lyu, C. E. Moore, T. Liu, Y. Yu, Y. Wu, *J. Am. Chem. Soc.* **143**, 12766–12776 (2021).
44. D. Luo *et al*., *Science* **360**, 1442–1446 (2018).
45. Q. Jiang *et al*., *Nature* **611**, 278–283 (2022).
46. M. Stolterfoht *et al*., *Adv. Mater.* **32**, 2000080 (2020).
47. M. Stolterfoht *et al*., *Nat. Energy* **3**, 847–854 (2018).
48. H. Zhang *et al*., *Nat. Commun.* **12**, 3383 (2021).
49. M. V. Khenkin *et al*., *Nat. Energy* **5**, 35–49 (2020).
50. J. D. Benck, B. A. Pinaud, Y. Gorlin, T. F. Jaramillo, *PLOS ONE* **9**, e107942 (2014).
51. R. A. Kerner, B. P. Rand, *ACS Appl. Energy Mater.* **2**, 6097–6101 (2019).
52. T. Moot *et al*., *ACS Energy Lett.* **6**, 2038–2047 (2021).

ACKNOWLEDGMENTS

Funding: This research was made possible by the US Department of the Navy, Office of Naval Research (grant N00014-20-1-2572). This work was supported in part by Ontario Research Fund: Research Excellence program (ORF7-Ministry of Research and Innovation, Ontario Research Fund: Research Excellence Round 7). This work was also supported under award OSR-CRG2020-4350.2. S.M.P., H.R.A., and K.R.G. gratefully acknowledge the US Department of Energy, Office of Science, Office of Basic Energy Sciences, and the EPSCoR program, under award DE-SC0018208 for XPS measurements. K.R.R., Z.C., J.T.P., C.M.R., and K.R.G. acknowledge funding from the NSF through Cooperative Agreement 1849213. Supercomputing resources on the Lipscomb High Performance Computing Cluster were provided by the University of Kentucky Information Technology Department and Center for Computational Sciences (CCS). M.W. acknowledges funding from the European Union's Horizon 2020 Research and Innovation program under Marie Skłodowska-Curie Grant Agreement 101026353. T.W. and A.A. acknowledge the support of the NSF under award ECCS-1936527. A.A. acknowledges partial support by the Office of Naval Research under award N00014-20-1-2573. This work made use of the NUFAB and Keck-II facilities of Northwestern University's NUANCE Center, which has received support from the SHyNE Resource (NSF ECCS-2025633), the IIN, and Northwestern's MRSEC program (NSF DMR-1720139).
Author contributions: S.M.P., M.W., K.R.G., M.G., and E.H.S. conceived the idea and proposed the experimental and model design. S.M.P. and M.W. fabricated all the devices and conducted the characterization. S.M.P and H.R.A performed XPS and UPS characterization and data analysis. J.X., K.R.R., and C.M.R. conducted the DFT simulation. Y.Y., C.L., and M.G.K. measured TOF-SIMS and fabricated PSMs. F.T.E., M.W., S.M.Z., and M.G. conducted the PL characterization and data analysis. K.D., T.W., and A.A. performed the GIWAXS measurements. S.T. and S.M.P. measured the TR spectra. L.G., B.C., H.C., L.Z., A.M., and Z.W. helped with the device fabrication and material characterization. Z.C. and J.T.P. performed the contact-angle measurements. M.W., S.M.P., J.X., K.R.G., M.G., and E.H.S. cowrote the manuscript. All authors contributed to data analysis, read, and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are provided in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adi4107
Materials and Methods
Supplementary Text
Figs. S1 to S31
Tables S1 to S4
References (*53*–*63*)

Submitted 24 April 2023; accepted 6 June 2023
10.1126/science.adi4107

WOOD ENGINEERING

Multiplex CRISPR editing of wood for sustainable fiber production

Daniel B. Sulis[1,2,3], Xiao Jiang[4], Chenmin Yang[1,2,3], Barbara M. Marques[2,3], Megan L. Matthews[5,6], Zachary Miller[4], Kai Lan[4], Carlos Cofre-Vega[2,3], Baoguang Liu[1,2,3,7], Runkun Sun[4], Henry Sederoff[2], Ryan G. Bing[8], Xiaoyan Sun[9], Cranos M. Williams[3,5], Hasan Jameel[4], Richard Phillips[4], Hou-min Chang[4], Ilona Peszlen[4], Yung-Yun Huang[10], Wei Li[11], Robert M. Kelly[8], Ronald R. Sederoff[1,2,11], Vincent L. Chiang[1,2,4,11], Rodolphe Barrangou[1,3,12]*, Jack P. Wang[1,2,3,11]*

The domestication of forest trees for a more sustainable fiber bioeconomy has long been hindered by the complexity and plasticity of lignin, a biopolymer in wood that is recalcitrant to chemical and enzymatic degradation. Here, we show that multiplex CRISPR editing enables precise woody feedstock design for combinatorial improvement of lignin composition and wood properties. By assessing every possible combination of 69,123 multigenic editing strategies for 21 lignin biosynthesis genes, we deduced seven different genome editing strategies targeting the concurrent alteration of up to six genes and produced 174 edited poplar variants. CRISPR editing increased the wood carbohydrate-to-lignin ratio up to 228% that of wild type, leading to more-efficient fiber pulping. The edited wood alleviates a major fiber-production bottleneck regardless of changes in tree growth rate and could bring unprecedented operational efficiencies, bioeconomic opportunities, and environmental benefits.

Wood is the most abundant carbon biomass on Earth and the major source of sustainable green fibers (*1*). Globally, ≈315 billion metric tons of carbon are stored as wood, representing ≈57% of the biogenic carbon sink (*1*). The biomass supplies >170 million metric tons of virgin fibers annually (*2*) to meet the growing demand for renewable tissue, paper, packaging, textile, and other fiber products, including structural materials (*3*). Despite the importance of wood fibers, their production has remained largely limited to undomesticated forest trees with often suboptimal wood properties that hamper production efficiency. The propensity for efficient isolation of cellulosic fibers from wood is largely determined by the content and composition of lignin (*4*–*6*), one of three major components of wood (*7*). Lignin is a phenolic polymer that cross-links with cellulose and hemicelluloses in the secondary cell walls of vascular plants (*8*, *9*). The polymer in angiosperm wood is formed by free radical polymerization of three major monolignol precursors, 4-coumaryl, coniferyl, and sinapyl alcohols, which form the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units in lignin (*10*–*12*). The monolignols are biosynthesized from phenylalanine through a series of enzymatic reactions in a metabolic grid consisting of at least 11 enzyme families and 24 metabolites (*13*). The pathway is mediated by hundreds of regulatory influences, encompassing transcriptional (*14*), translational (*15*), and posttranslational regulations (*16*, *17*); enzyme–enzyme interactions (*18*–*20*); and metabolic regulations (*21*, *22*). More than five decades of research have extensively investigated the individual components of lignin biosynthesis and determined the effects of their perturbation on lignin content and composition in diverse plant species (*6*, *13*, *23*). However, these efforts have predominantly focused on the modification of single genes or gene families, whereas the combinatorial effects of multigenic perturbations have remained elusive. Here, we show that strategic multiplex CRISPR editing of monolignol biosynthesis genes improves wood properties beyond what can be achieved by editing single genes or gene families and debottlenecks a key operational constraint in industrial pulp mills. These improvements substantially increase fiber production capacity while reducing the global warming potential of pulp mills, leading to a more sustainable and efficient fiber bioeconomy.

[1]TreeCo, Raleigh, NC, USA. [2]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA. [3]NC Plant Sciences Initiative, North Carolina State University, Raleigh, NC, USA. [4]Department of Forest Biomaterials, North Carolina State University, Raleigh, NC, USA. [5]Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA. [6]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Champaign, IL, USA. [7]Department of Forestry, Beihua University, Jilin, China. [8]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA. [9]Molecular Education, Technology and Research Innovation Center (METRIC), North Carolina State University, Raleigh, NC, USA. [10]Department of Operations Research, North Carolina State University, Raleigh, NC, USA. [11]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China. [12]Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, USA.
*Corresponding author. Email: rbarran@ncsu.edu (R.B.); jpwang@ncsu.edu (J.P.W.)

Multiplex genome editing strategies for concurrent wood trait improvement

Gene targets for multiplex genome editing were identified using our established predictive model for monolignol biosynthesis (*13*, *15*, *22*, *24*). The model predicts the transduction of quantitative relationships from gene transcript abundances to absolute enzyme abundances, pathway metabolic fluxes, and 25 wood chemical and physical properties. Using the predictive model, we explored 69,123 sets of multigenic editing strategies to reveal the extent to which individual wood properties can be modified through multiplex editing of monolignol genes (Fig. 1, A and B). The 69,123 strategies encompass 48,831 gene combinations that represent all possible permutations of loss-of-function editing of up to six genes, and 20,292 gene combinations that target the combinatorial editing of up to three genes and/or overexpression of one gene (data S1A). For each gene combination, the model estimated the corresponding abundances of monolignol enzymes, pathway metabolic fluxes, and 25 wood properties. Targeting multiple genes expanded the range of phenotypic variation attainable by genome editing compared with single-gene edits (Fig. 1C and data S1B). For example, multigenic editing could reduce lignin content to 50.7% of the wild-type level, whereas single-gene edits only reduced lignin content to 61.3% of the wild-type level (Fig. 1C and data S1B).

Wood with low lignin content and a high syringyl-to-guaiacyl (S/G) ratio is ideal for fiber production (*25*). We mined all 69,123 strategies to identify gene combinations that are predicted to reduce lignin content by at least 15% and increase the carbohydrate-to-lignin (C/L) ratio by at least 200% and that have a higher S/G ratio than wild type. The C/L ratio is an indicator of the potential maximum cellulosic yield for wood fiber (*13*). The strategies must also have predicted growth characteristics (e.g., tree height) that are comparable to (>75%) or exceed those of wild-type controls. Of the 69,123 strategies, only 347 (0.5%) matched the aforementioned criteria in lignin content, C/L and S/G ratios, and tree growth (Fig. 1B and data S1C), highlighting the need to mine such strategies to practically test the most-promising combinations. All 347 strategies target at least two monolignol genes, and 99.7% of the strategies target at least three genes (Fig. 1D and data S1C). The number of target genes affirms the need for a multigenic approach to improve fiber traits in *Populus trichocarpa*. Lignin gene families that most frequently appeared in these 347 strategies were *C3H* (coumarate 3-hydroxylase), *CCoAOMT* (caffeoyl-CoA O-methyltransferase), *AldOMT* (aldehyde *O*-methyltransferase), *PAL* (phenylalanine ammonia-lyase), *C4H* (cinnamate 4-hydroxylase), and *CAD* (cinnamyl alcohol

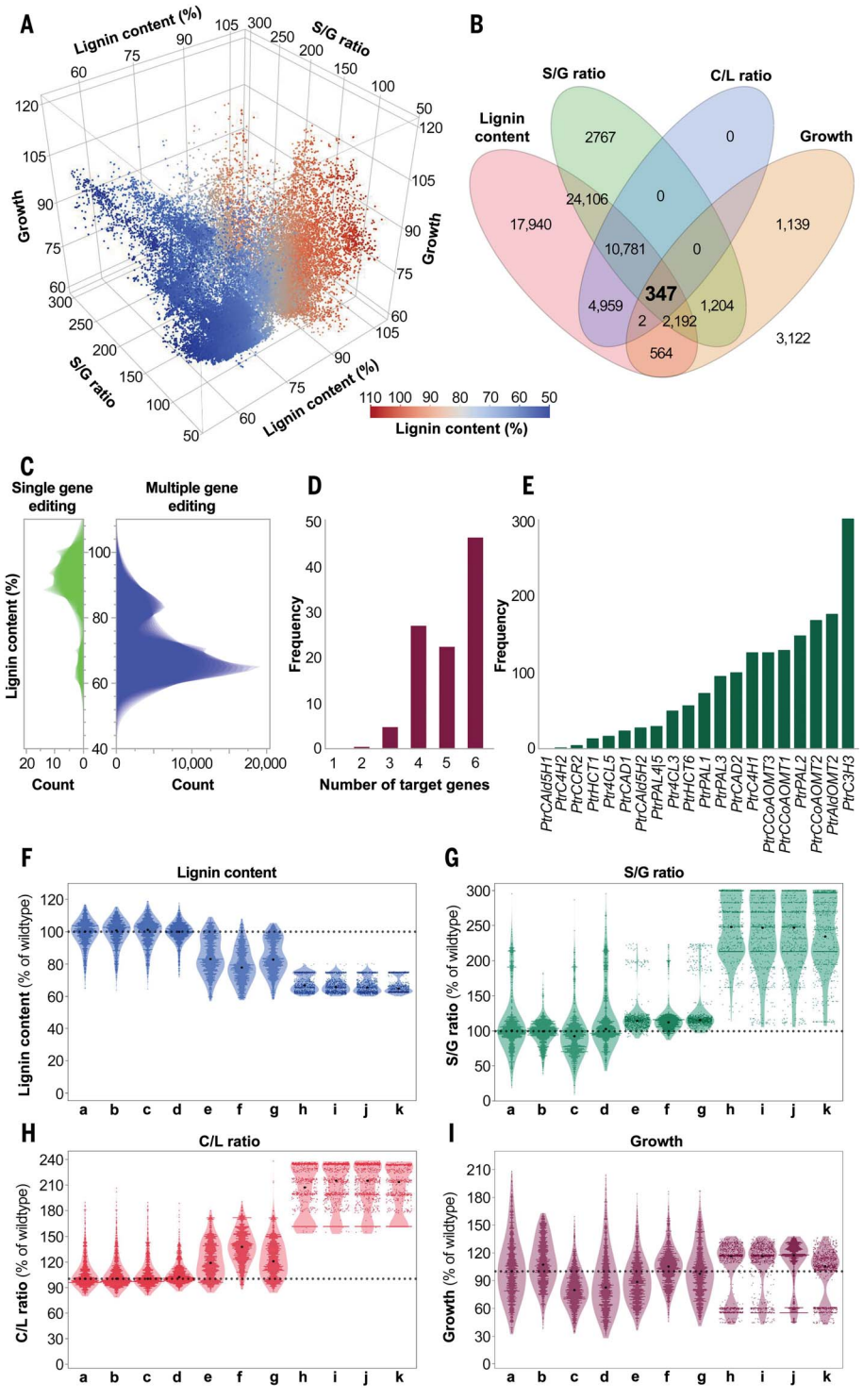


Fig. 1. Identification of multiplex CRISPR editing strategies to improve wood fiber traits. (**A**) Three-dimensional scatterplot of predicted lignin content, S/G ratio, and tree growth (stem diameter) for 69,123 genome editing strategies. (**B**) Venn diagram of the 69,123 strategies showing the relationships between predicted lignin content, S/G ratio, C/L ratio, and tree growth (height). (**C**) Distribution of lignin content in single-gene and multigene editing strategies. (**D**) The total number of target genes in the selected 347 strategies. (**E**) The frequency of each monolignol gene in the 347 strategies. (**F** to **I**) Global sensitivity analysis of wild-type (a), single-gene editing (b to d), and multigene editing (e to k) strategies for lignin content (F), S/G ratio (G), C/L ratio (H), and tree growth (height) (I). For additional information, see fig. S1.

dehydrogenase) (Fig. 1E). We then performed sensitivity analysis for these gene families to examine the robustness of fiber trait improvements in response to varying transcript abundances of nontargeted monolignol genes (Fig. 1, F to I; fig. S1; and data S2 and S3) and selected seven strategies for editing in *P. trichocarpa*. The seven strategies (table S1) encompassed various numbers of target genes (from three to six) and were selected on the basis of the extent and robustness of predicted improvement in fiber traits (reduced lignin, increased S/G and C/L ratios, and good growth) (data S2 and S3). The selected strategies showed a predicted reduction in lignin content of up to 35% compared with wild type (Fig. 1F and data S3). S/G and C/L ratios increased up to 248 and 215%, respectively (Fig. 1, G and H, and data S3), with no change in tree height (Fig. 1I and data S3).

Generating genetic diversity in poplar using CRISPR

To test the seven strategies in planta for their modulation of wood properties for fiber production, a multiplex CRISPR construct (*26*) was assembled for each strategy and delivered into *P. trichocarpa* using *Agrobacterium tumefaciens* (*27*) (figs. S2 to S4, data S4, and tables S2 and S3). We generated 174 independent lines of multiplex CRISPR–edited *P. trichocarpa*, including 78 lines that target the concurrent editing of three monolignol genes, 20 lines that target four genes, 41 lines that target five genes, and 35 lines that target six genes. The edited lines exhibited varying extents of loss-of-function mutations of the target genes (Fig. 2A, fig. S5, and data S5). Biallelic loss-of-function editing of all target genes was obtained in *P. trichocarpa* for strategies that target the concurrent editing of three or four genes (Fig. 2A, fig. S5, and data S5). For strategies that target five or six genes, none of the 76 edited lines showed complete loss-of-function editing of both alleles in all target genes. Nonetheless, we identified several lines (e.g., J-7 and K-28) that harbor substantial editing of most target genes (fig. S5 and data S5). The frequency of biallelic edits varied between target genes, ranging from 86.1% for *PtrPAL4* and *PtrPAL5* to 9.3% for *PtrCCoAOMT2* (Fig. 2B). Most edits consisted of small insertions and deletions (INDELs) of 1 to 3 nucleotides (nt) in the immediate vicinity of the targeted cleavage site (Fig. 2, C and D, and fig. S6). However, deletions of up to 19 nt were also occasionally observed (Fig. 2D). No off-target edits were detected in our edited lines (table S4). We used RNA sequencing to evaluate two potential xylem-expressing off-target genomic loci in 12 edited *P. trichocarpa* lines (data S6). The absence of detectable off-target editing suggests that genome editing is highly specific in *P. trichocarpa*. The broad

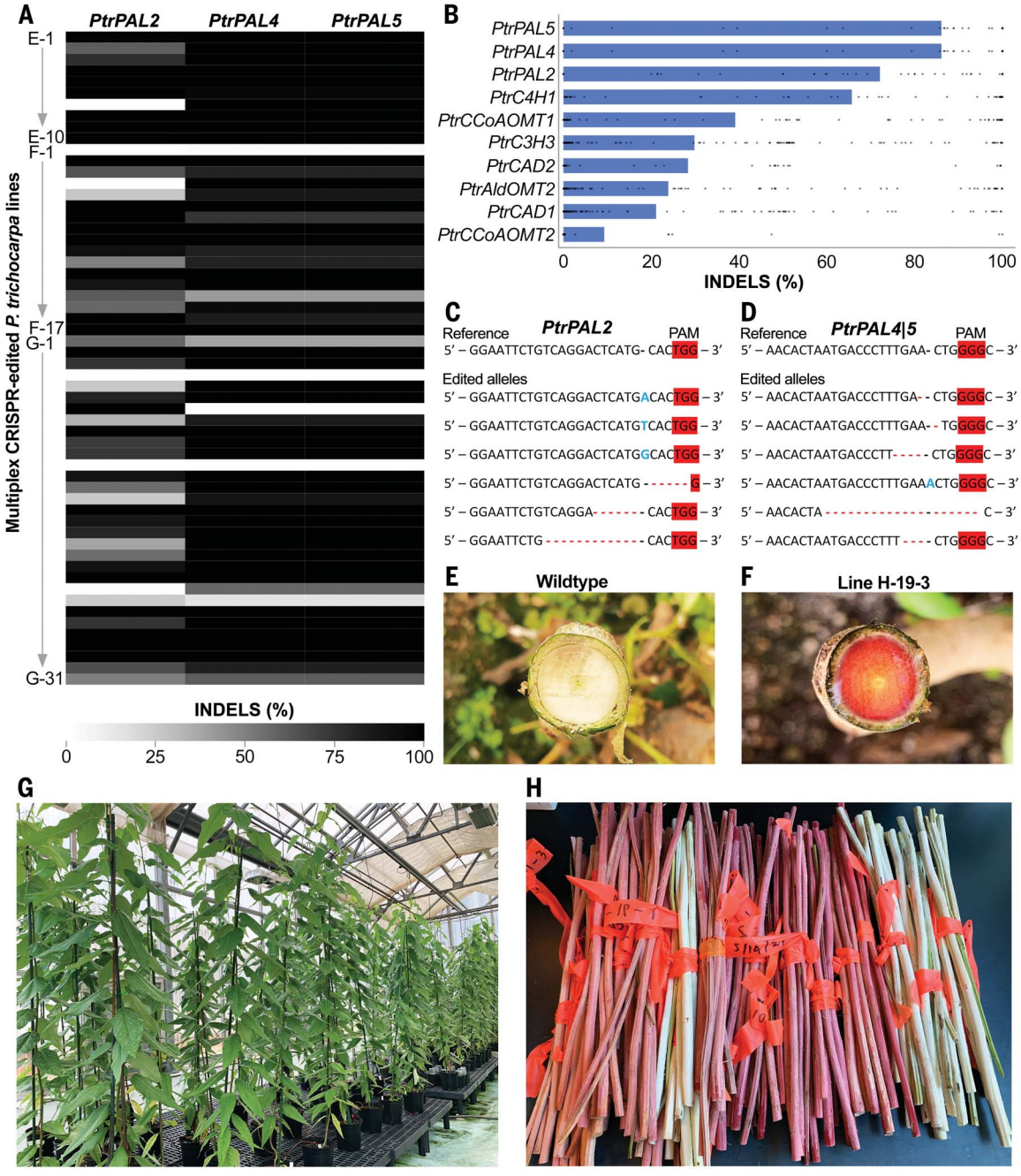


Fig. 2. Genetic diversity in multiplex CRISPR–edited *P. trichocarpa*. (**A**) Percentage INDELs of each target gene in the 58 CRISPR-edited *P. trichocarpa* lines targeting the concurrent editing of three monolignol genes (*PtrPAL2*, *PtrPAL4*, and *PtrPAL5*). (**B**) Average percentage INDELs of the target monolignol genes in the 174 edited *P. trichocarpa* lines. (**C** and **D**) Most-frequent CRISPR-Cas–mediated mutations in *PtrPAL2* (C) and *PtrPAL4* and *PtrPAL5* (D). Stem cross sections of wild-type (**E**) and CRISPR-edited H-19-3 trees (**F**), with the H-19-3 tree showing distinct red coloration of the xylem. (**G**) Six-month-old greenhouse-grown CRISPR-edited and wild-type *P. trichocarpa*. (**H**) Harvested stem segments from CRISPR-edited and wild-type *P. trichocarpa*.

variation in the editing profiles of target genes in our 174 edited lines created genetic diversity in monolignol biosynthesis not present in nature (Fig. 2, E to H). Such genetic diversity enables an exploration of how multigenic pathway perturbations can combinatorially regulate wood formation and wood utilization, as well as enhance genetic diversity.

Phenotypic characterization of multiplex CRISPR–edited poplar

Multiplex editing of monolignol genes in poplar substantially altered the chemical and physical properties of the wood (Fig. 3, A to D; fig. S7; and data S7 to S10), as predicted by the lignin model (fig. S8 and table S5). Stem wood of 6-month-old wild-type *P. trichocarpa* contains 20.9% lignin and has a C/L ratio of 3.0 (data S7). In the multiplex-edited wood, lignin content is reduced by up to 49.1%, and the C/L ratio increased up to 228%, compared with wild type (Fig. 3, E and F, and data S7, determined as averages across 65 lines). Two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) (Fig. 3, G to J, and figs. S9 to S11)

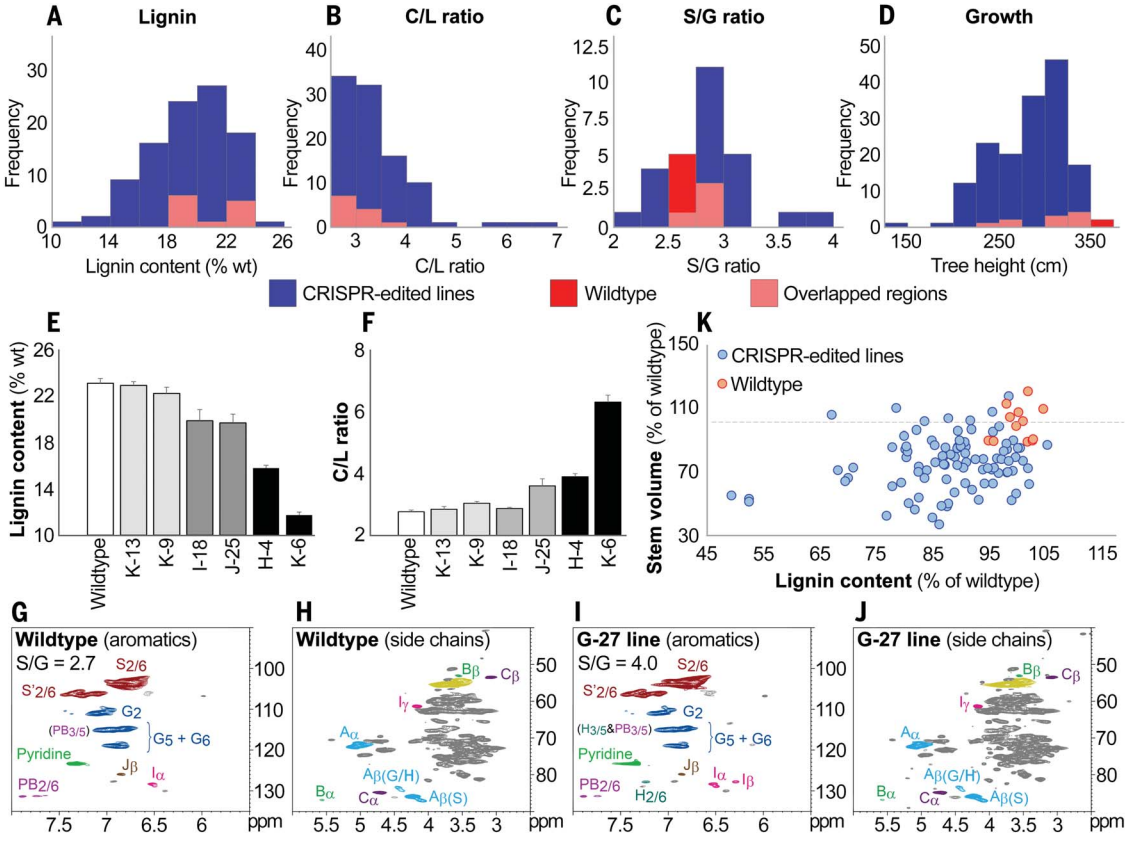


Fig. 3. Phenotypic variation in CRISPR-edited and wild-type *P. trichocarpa*. (**A** to **D**) Distribution of lignin content (A), C/L ratio (B), S/G ratio (C), and tree growth (height) (D) in CRISPR-edited and wild-type *P. trichocarpa*. Blue and red colors stand for CRISPR-edited lines and wild-type control measurements, respectively. Pink represents the overlapped regions of the blue (CRISPR-edited lines) and red (wild-type controls) bars. Lignin, C/L ratio, and growth data are from E, F, G, H, I, J, and K constructs. S/G ratio data are from E, F, and G constructs. Average lignin content (**E**) and C/L ratio (**F**) of selected CRISPR-edited lines with improved fiber traits. Error bars represent one standard error of up to three biological replicates. (**G** to **J**) 2D NMR ^{13}C-^{1}H (heteronuclear single quantum coherence) correlation spectra (side-chain and aromatic regions) of wild type [(G) and (H)] and the G-27 edited line (see fig. S8 for other edited lines). Contours in these regions were used to estimate the distribution of lignin interunit linkages and lignin composition, namely S/G/H ratios, as well as *p*-hydroxybenzoate (PB). ppm, parts per million. (**K**) Quantitative relationship between lignin content and wood volume in CRISPR-edited and wild-type *P. trichocarpa*.

revealed that multigenic strategies also modulated lignin composition, increasing the S/G ratio from 2.7 in wild type to as high as 4.0 in the edited lines (Fig. 3C and data S8). As expected, lines harboring single-gene edits showed either no reduction or a mild reduction in lignin content (0.1 to 12.6%) (e.g., lines G-7, I-17, and K-9; fig. S12 and data S5 and S7). The most significant lignin reductions were observed in edited trees that harbor four to six gene edits, but strategies that targeted three genes also showed significant lignin reductions of up to 32% (Fig. 3E, fig. S12, and data S7). Edited lines with significant changes in wood chemical properties predominantly harbored a single-allele loss-of-function mutation in *PtrC3H3* combined with the mutation of either one or both alleles of *PtrAldOMT2* and *PtrCAD1* (fig. S12). Wood elasticity was not significantly different between wild-type and genome-edited *P. trichocarpa*, except for two lines (E-9 and G-25), which had reduced wood elasticity (fig. S7 and data S9). No change in wood density greater than 15.1% (detection limit due to the large technical and biological variation; see materials and methods in the supplementary materials) of the wild-type level was observed in the edited *P. trichocarpa* (data S9). Several CRISPR-edited lines with significant reductions in lignin content showed a reduction in tree growth (e.g., I-1, J-25, and K-6) (Fig. 3, D and K; figs. S7 and S12; and data S7 and S10). Some multiplex genome editing strategies can circumvent certain shortcomings inherent to typical single-knockout strategies that often have undesirable phenotypic effects on combinations of tree characteristics and wood properties. Moving forward, it will be necessary to monitor the impact of the genome editing outcomes on the tree phenotypic properties and traits relevant to the industry in long-term field trials.

Projected industrial benefits of debottlenecking fiber production

Kraft pulping is the dominant process for industrial wood fiber production (*28*–*30*). To understand the techno-economic impacts of CRISPR-edited wood on kraft pulping, we used a Carolina Pulp and Paper (CPP) mill model (fig. S13) based on an existing industrial Brazilian pulp mill with an annual production of 1.24 million tons of pulp. Pulp yield and production efficiency may be significantly increased by multiplex editing of monolignol genes (Fig. 4, A and B), as evidenced by micropulping wood with varying lignin content and composition. Reducing lignin content proportionally increases pulp yield and could reduce the usage of pulping chemicals (Fig. 4, A and B; figs. S14 to S17; and tables S6 to S8) that are undesirable from an environmental standpoint. Furthermore, low-lignin wood could reduce the solid content of black liquor,

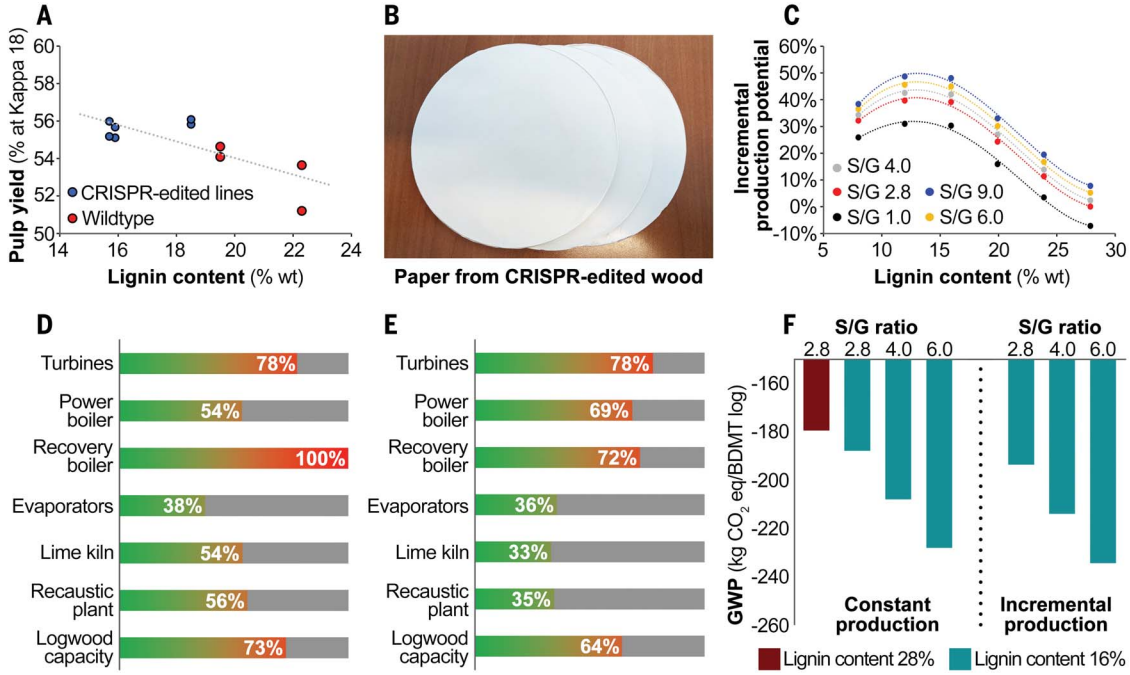


Fig. 4. Techno-economic analysis and global warming potential of CRISPR-edited wood in industrial kraft pulping. (**A**) Correlation between lignin content and pulp yield in CRISPR-edited and wild-type *P. trichocarpa*. (**B**) Paper disks produced from CRISPR-edited *P. trichocarpa*. (**C**) The impact of varying lignin content and S/G ratio on the incremental production potential of an industrial kraft pulp mill. (**D** and **E**) Operational capacity of different components of an industrial kraft pulp mill using wood with 28% lignin (D) or 16% lignin (E) at an S/G ratio of 2.8. (**F**) Total life-cycle GWP of producing pulp from 1 BDMT log. The scenarios include varied S/G ratios (2.8, 4.0, or 6.0), lignin content (28 or 16%), and production modes (constant or incremental).

thereby debottlenecking the recovery boiler, arguably the most crucial and rate-limiting energetic component of pulp mills (Fig. 4, C to E, and fig. S18). The debottlenecking then enables an incremental production potential of the pulp mill by up to 40% (Fig. 4, C to E, and fig. S18). Additionally, increasing the C/L ratio in wood means that less biomass is required to provide the same amount of cellulose (fig. S17). The benefits of increasing the S/G ratio encompassed the reduction of pulping chemicals and energy savings on chemical recovery (figs. S16 and S18). Tremendous financial benefits may be obtained by reducing lignin from 28% to 16% and by increasing the S/G ratio from 2.8 to 4.0 (figs. S19 to S21). Such changes are predicted to increase the net present value (at a 14% hurdle rate) from −$359.3 million to $1047.1 million USD for pulp mills using natural gas as supplementary energy, and from $269.2 million to $1856.9 million USD for mills using biomass as supplementary energy (figs. S19, C and D, S20, and S21). Notably, some of these benefits are achieved even in cases of biomass reduction because of industrial operational efficiency gains (figs. S20 and S21). Although the magnitude of the financial gains is forecasted, accuracy will eventually be determined by the physiological attributes of the harvested trees at industrial scale.

Sustainability impact of optimized fiber production

The pulp and paper industry is a major contributor to greenhouse gas emissions. The US Environmental Protection Agency reported 35 million metric tons of CO_2 equivalent emissions from the pulp and paper industry in the US in 2020 (*31*). Globally, the annual direct CO_2 emissions from pulp and paper production reached 168 million metric tons (*32*). We assessed the impacts of CRISPR-edited wood on the carbon footprint of pulp production. Life-cycle global warming potential (GWP) was estimated for pulp production from 1 bone dry metric ton (BDMT) of CRISPR-edited wood and compared with that of wild-type wood. For a pulp mill powered by natural gas, reducing lignin content from 28% to 16% decreases the GWP by up to 8% (Fig. 4F, figs. S22 to S26, and tables S9 and S10), and increasing the S/G ratio from 2.8 to 4.0 leads to a further 10% reduction in the GWP per BDMT of logs processed (Fig. 4F, figs. S22 to S26, and tables S9 and S10). Altogether, incremental pulp production using edited wood with reduced lignin and an increased S/G ratio could reduce GWP by up to 20% (Fig. 4F, figs. S22 to S26, and tables S9 and S10), providing substantial benefits to environmental conservation and climate change mitigation. These results provide insights into the potential of genome editing technologies for tree improvement, which may be further corroborated in field trials in relevant environmental conditions. This study illustrates how strategic multiplex editing to alter wood composition could enable more-sustainable fiber production with remarkable operational efficiencies, bioeconomic value creation, and tangible environmental benefits.

REFERENCES AND NOTES

1. Y. M. Bar-On, R. Phillips, R. Milo, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6506–6511 (2018).
2. RISI, Fastmarkets, "RISI annual review of global pulp and paper statistics" (2022); https://www.risiinfo.com/product/annual-review-of-global-pulp-paper-statistics.
3. S. Xiao *et al.*, *Science* **374**, 465–471 (2021).
4. J. Ralph *et al.*, *Science* **277**, 235–239 (1997).
5. R. Vanholme *et al.*, *Science* **341**, 1103–1106 (2013).
6. F. Chen, R. A. Dixon, *Nat. Biotechnol.* **25**, 759–761 (2007).
7. R. C. Pettersen, *Adv. Chem. Ser.* **207**, 57–126 (1984).
8. Q. Li *et al.*, *Plant Biotechnol. J.* **12**, 1174–1192 (2014).
9. K. V. Sarkanen, C. H. Ludwig, *Lignins: Occurrence, Formation, Structure and Reactions* (Wiley-Interscience, 1971).
10. W. Boerjan, J. Ralph, M. Baucher, *Annu. Rev. Plant Biol.* **54**, 519–546 (2003).
11. R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan, *Plant Physiol.* **153**, 895–905 (2010).
12. K. Freudenberg, *Science* **148**, 595–600 (1965).
13. J. P. Wang *et al.*, *Nat. Commun.* **9**, 1579 (2018).
14. H. Chen *et al.*, *Plant Cell* **31**, 602–626 (2019).
15. M. L. Matthews, J. P. Wang, R. Sederoff, V. L. Chiang, C. M. Williams, *PLOS Comput. Biol.* **16**, e1007197 (2020).
16. J. P. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8481–8486 (2015).

17. D. B. Sulis, J. P. Wang, *Front. Plant Sci.* **11**, 914 (2020).
18. J. P. Wang, B. Liu, Y. Sun, V. L. Chiang, R. R. Sederoff, *Front. Plant Sci.* **9**, 1942 (2019).
19. H. C. Chen *et al*., *Plant Cell* **26**, 876–893 (2014).
20. H.-C. Chen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 21253–21258 (2011).
21. J. P. Wang *et al*., *Plant Cell* **26**, 894–914 (2014).
22. J. P. Wang *et al*., *Curr. Opin. Biotechnol.* **56**, 187–192 (2019).
23. A. J. Ragauskas *et al*., *Science* **344**, 1246843 (2014).
24. M. L. Matthews, J. P. Wang, R. Sederoff, V. L. Chiang, C. M. Williams, *Comput. Struct. Biotechnol. J.* **19**, 168–182 (2020).
25. R. B. Santos, E. A. Capanema, M. Y. Balakshin, H.-M. Chang, H. Jameel, *BioResources* **6**, 3623–3637 (2011).
26. R. Ueta *et al*., *Sci. Rep.* **7**, 507 (2017).
27. J. Song, S. Lu, Z. Z. Chen, R. Lourenco, V. L. Chiang, *Plant Cell Physiol.* **47**, 1582–1589 (2006).
28. P. J. Kleppe, *Tappi* **53**, 35–47 (1970).
29. J. Gierer, *Wood Sci. Technol.* **14**, 241–266 (1980).
30. N. Bijok, J. Fiskari, R. R. Gustafson, V. Alopaeus, *Chem. Eng. J.* **438**, 135548 (2022).
31. US Environmental Protection Agency, Greenhouse Gas Reporting Program (GHGRP) (2021); https://www.epa.gov/ghgreporting/ghgrp-pulp-and-paper#map-facilities.
32. International Energy Agency, Tracking Industry (2020); https://www.iea.org/reports/tracking-industry-2020 [accessed 18 July 2021].

ACKNOWLEDGMENTS

We thank K. Osakabe for providing the multiplex genome editing vectors and associated protocols for constructs assembly used in this paper. **Funding:** This work was supported by National Institute of Food and Agriculture of the US Department of Agriculture–Agriculture and Food Research Initiative grant 2018-67021-27716; North Carolina State University Chancellor's Innovation Fund grant 190549MA; North Carolina State University Goodnight Early Career Innovator Award (to J.P.W.); US National Science Foundation Small Business Technology Transfer Program grant 2044721; Cooperative State Research Service of the US Department of Agriculture grant NCZ04214; and North Carolina Specialty Crop Block Grants 19-019-4018, 19-092-4012, and 20-070-4013. **Author contributions:** Conceptualization: J.P.W., V.L.C., R.R.S., and R.B. Methodology: D.B.S., X.J., C.Y., M.L.M., C.M.W., H.J., R.P., H.-m.C., I.P., Y.-Y.H., and W.L. Investigation: D.B.S., X.J., C.Y., B.M.M., Z.M., K.L., C.-C.V., B.L., R.S., H.S., R.G.B., and X.S. Funding acquisition: J.P.W., R.M.K., and R.B. Supervision: J.P.W., R.B., C.M.W., V.L.C., H.J., R.P., H.-m.C., I.P., and R.R.S. Writing: D.B.S., J.P.W., X.J., K.L., B.M.M., C.-C.V., and R.B. **Competing interests:** R.B. and J.P.W. are shareholders of TreeCo. All other authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add4514
Materials and Methods
Supplementary Text
Figs. S1 to S26
Tables S1 to S10
References (*33*–*40*)
Data S1 to S10
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 12 June 2022; resubmitted 20 November 2022
Accepted 16 May 2023
10.1126/science.add4514

QUANTUM NETWORKS

Multichip multidimensional quantum networks with entanglement retrievability

Yun Zheng[1]†, Chonghao Zhai[1]†, Dajian Liu[2,3]†, Jun Mao[1], Xiaojiong Chen[1], Tianxiang Dai[1], Jieshan Huang[1], Jueming Bao[1], Zhaorong Fu[1], Yeyu Tong[4], Xuetong Zhou[5], Yan Yang[6], Bo Tang[6], Zhihua Li[6], Yan Li[1,7,8,9], Qihuang Gong[1,7,8,9], Hon Ki Tsang[5], Daoxin Dai[2,3]*, Jianwei Wang[1,7,8,9]*

Quantum networks provide the framework for quantum communication, clock synchronization, distributed quantum computing, and sensing. Implementing large-scale and practical quantum networks relies on the development of scalable architecture and integrated hardware that can coherently interconnect many remote quantum nodes by sharing multidimensional entanglement through complex-medium quantum channels. We demonstrate a multichip multidimensional quantum entanglement network based on mass-manufacturable integrated-nanophotonic quantum node chips fabricated on a silicon wafer by means of complementary metal-oxide-semiconductor processes. Using hybrid multiplexing, we show that multiple multidimensional entangled states can be distributed across multiple chips connected by few-mode fibers. We developed a technique that can efficiently retrieve multidimensional entanglement in complex-medium quantum channels, which is important for practical uses. Our work demonstrates the enabling capabilities of realizing large-scale practical chip-based quantum entanglement networks.

Quantum networks aim to transmit, store, and process quantum information in a mesh of quantum nodes coherently linked by quantum channels, going beyond the capabilities of classical internet (*1*, *2*). It promises many applications, such as quantum key distribution (QKD) for secure communications, higher-precision clock synchronization, distributed quantum metrology, and distributed quantum computation. The key to upgrading current QKD networks (*3*–*5*) is realizing entanglement networks that can robustly share complex entangled states among a large number of remote quantum nodes (*1*, *2*).

Realizing large-scale and practical quantum entanglement networks remains experimentally challenging. It requires scalable quantum hardware, technologies, and architecture for the distributions of entangled states among a large number of quantum nodes through high-capacity quantum channels. Quantum photonic integrated circuits are an emerging hardware system for quantum information processing and communication (*6*–*8*). U… integrated photonic chips, to date, point-… point QKD (*9*–*13*) and entanglement distribution (*14*, *15*) have been demonstrated in a single chip–to–single chip system, and single microring resonators have been used as photon sources for QKD and networks (*16*–*18*). Comprehensive integrated quantum nodes in networks require full-spectrum monolithic integration of devices for encoding, decoding, multiplexing, operation, and detection of quantum states of light. In contrast to single-chip or single chip–to–single chip quantum experiments, realizing multichip quantum networks remains challenging because it not only requires the monolithic integration of different devices but requires high

[1]State Key Laboratory for Mesoscopic Physics, School of Physics, Peking University, Beijing 100871, China. [2]State Key Laboratory for Extreme Photonics and Instrumentation, College of Optical Science and Engineering, Ningbo Research Institute, International Research Center for Advanced Photonics, ZJU–Hangzhou Global Scientific and Technological Innovation Center, Zhejiang University, Hangzhou 310058, China. [3]Intelligent Optics and Photonics Research Center, Jiaxing Research Institute, Zhejiang University, Jiaxing 314000, China. [4]Microelectronics Thrust, Function Hub, The Hong Kong University of Science and Technology (Guangzhou), China. [5]Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, 999077 Hong Kong. [6]Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China. [7]Frontiers Science Center for Nano-optoelectronics and Collaborative Innovation Center of Quantum Matter, Peking University, Beijing 100871, China. [8]Peking University Yangtze Delta Institute of Optoelectronics, Nantong 226010, Jiangsu, China. [9]Hefei National Laboratory, Hefei 230088, China.
***Corresponding author. Email: dxdai@zju.edu.cn (D.D.); jww@pku.edu.cn (J.W.)**
†These authors contributed equally to this work.

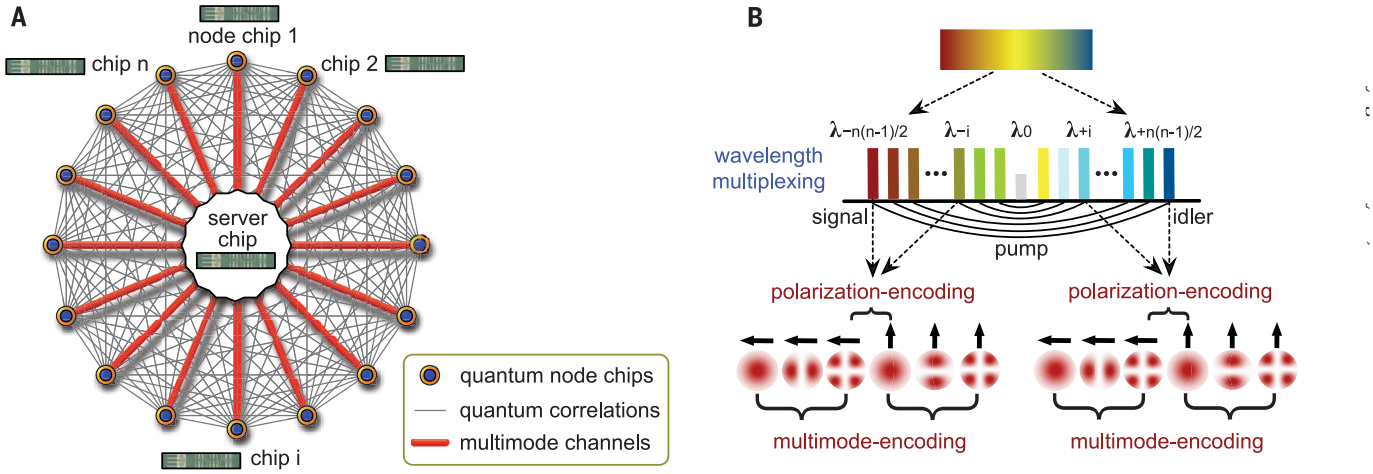


Fig. 1. Architecture for multichip multidimensional entanglement networks. (**A**) Graphic representation of a fully connected entanglement network. It consists of one central server chip, *n* quantum node chips (vertices), *n* multimode quantum channels (red edges), and quantum correlations (gray edges). All node chips can be fully connected by sharing $n(n-1)/2$ pairs of *d*-dimensional entangled photons with different wavelengths, initially produced on the server chip. Each multimode channel carries $n-1$ wavelength-multiplexed single photons, to be received and processed at one node chip. In the experiment, we prototyped an entanglement network having $n = 3$ and $d = 4$. (**B**) Hybrid encoding and multiplexing of entangled photons. The entangled photons are wavelength-demultiplexed from a broadband source and then remultiplexed for entanglement distributions through MMFs between node chips. At each wavelength, the *d*-dimensional quantum states are path-encoded on chip, which are coherently transformed to hybrid polarization mode–encoding states in fibers.

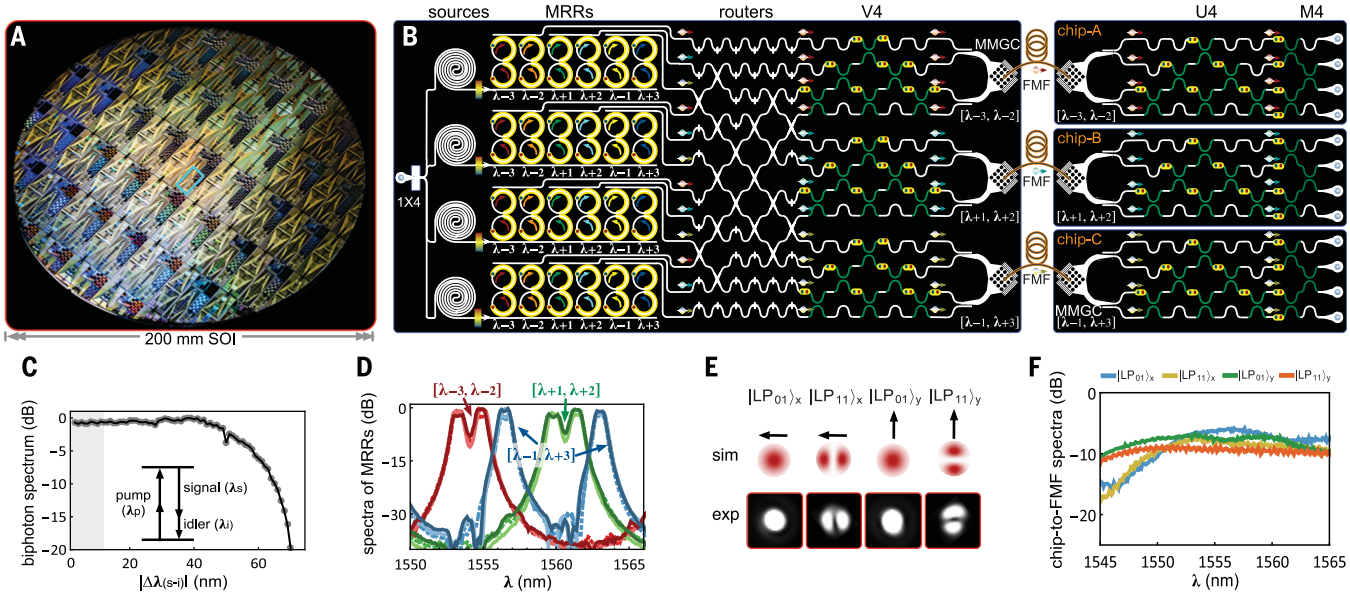


Fig. 2. Integrated silicon-photonic quantum networking chips. (**A**) Photograph for a 200-mm silicon-on-insulator (SOI) wafer that contains 27 copies of networking chips. The cyan box indicates a single copy, including a server chip and a node chip. (**B**) Circuit diagram. The network consists of a server chip that produces three pairs of four-dimensional entangled photons and three node chips (A, B, and C) connected by three separate FMFs. The triangular linear optical circuits of V_4 and U_4 are fully programmable to implement arbitrary unitary operations for the QER algorithm. M_4 on the node chips allows arbitrary projective measurement. Green parts indicate reconfigurable beamsplitters, and yellow parts indicate thermo-optic phase shifters. (**C**) Measured biphoton spectrum of integrated sources. The grayed area indicates the wavelength window in this experiment. (Inset) An energy diagram for the nondegenerate SFWM process. (**D**) Measured spectra of coupled MRRs. Four sets of coupled MRRs (12 in total, colored) were measured, showing high-level spectral overlapping and low channel cross-talk. (**E** and **F**) Characterizations of the chip-to-fiber system. The LP modes were selectively excited and measured. (E) Simulated and measured LP mode distributions at the FMF output facet. (F) Measured spectral response for these modes. The spectral response for one-dimensional grating couplers (used for light injection) has been subtracted.

quantum indistinguishability between different chips for scalable quantum networking. Meanwhile, practical quantum networking requires high-capacity and noise-resilient transmission of entanglement. Multidimensional quantum systems are known as one of the excellent candidates, which however are susceptible to environmental perturbations when traversing through complex media such as multimode fiber (MMF) or air-scattering channels (*19*, *20*). As a result, state scrambling and entanglement degeneration inevitably take place. Previous works have demonstrated the capabilities of unscrambling classical (*21*, *22*) and quantum (*23*, *24*) light in complex media by using a measure-and-inverse technique, which requires the reconstruction of transmission matrix and then the inverse operation to recover the states. Another all-optics unscrambling method has been recently proposed (*25*) and demonstrated for unitary transmission of classical light (*26*). Efficiently retrieving quantum states of light in complex media is of practical interest in real-world quantum networking systems. Moreover, at the architecture level, quantum networking devices and technologies are expected to sustain a large number of quantum nodes and high-capacity quantum channels and to be compatible with those well-developed ones in classical telecommunication networks, such as multiplexing devices and technologies.

In this work, we demonstrated multichip multidimensional entanglement networks. Our quantum networking chips were fabricated on silicon nanophotonic quantum circuits by using mass-manufacturable technologies. Three pairs of multidimensional entangled photons were generated on a server chip and coherently distributed across three quantum node chips by few-mode fibers (FMFs). We developed a quantum entanglement retrieving (QER) method to overcome the nonunitary mode scrambling and entanglement degradation, efficiently recovering multidimensional entangled states distributed across the multichip networks.

Networking architecture and hybrid multiplexing

The network is enabled by hybrid encoding and multiplexing of the photon's wavelength, path, mode, and polarization states (Fig. 1A). The network can be architected by a complete graph with n vertices (quantum node chips) and $n(n-1)/2$ edges (quantum correlations). It can be physically realized as follows: A server chip simultaneously produces $n(n-1)/2$ pairs of d-dimensional entangled photons ($|\Phi_d^+\rangle = 1/\sqrt{d}\sum_k |k\rangle|k\rangle$, $k = 0, 1, ..., d-1$) at different wavelengths [$\lambda_{\pm i}$, $i = 1, 2, ..., n(n-1)/2$] and shares them through n MMF quantum channels. In each channel, $n-1$ sets of single photons are wavelength-multiplexed to form quantum correlations (*27*, *28*) between paired chips, and each photon with the dimension of d is hybridly polarization-mode encoded (Fig. 1B). The simultaneous distribution of entanglement in the network is enabled through wavelength-multiplexing parallelism, and the transmission of multidimensional entanglement between chips is enabled by coherently converting the on-chip multipath-encoding to the linearly polarized (LP) multimode encoding with different polarizations in fibers. The network architecture can be highly scalable; it only requires wavelength-multiplexed or -demultiplexed bipartite entangled states rather than multipartite entangled states. Resources scale up linearly, requiring n node chips linked by n channels and d photon sources on the server chip. The upper bounds of nodes (n) and capacity (d) are determined by the spectral and modal bandwidth of the whole system, respectively.

The 200-mm wafer that contains at least 27 copies of identical quantum networking chips

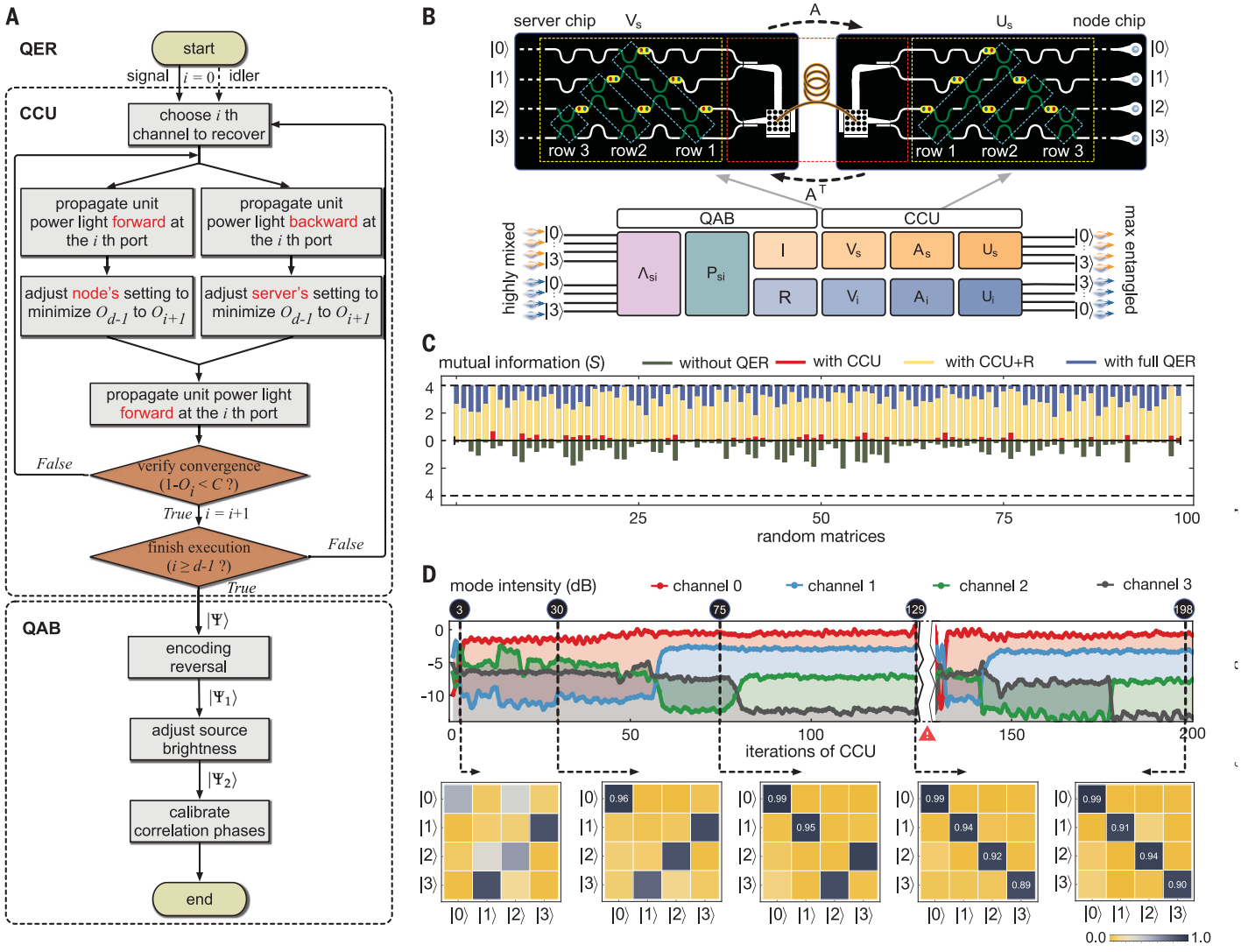


Fig. 3. Retrieving multidimensional entanglement in a multimode chip-fiber-chip system. (**A**) Flowchart of the QER algorithm consisting of two stages, CCU and QAB. The CCU is carried out to recover the mode mixing, and the QAB then retrieves maximal entanglement. O_i denotes the measured optical intensity (normalized) at the ith port. $\mathcal{C}$ is the convergent criterion; when the cross-talk $(1 - O_i)$ to other channels is less than $\mathcal{C}$, the CCU converges. (**B**) Implementing QER on the two reprogrammable photonic chips, linked by a FMF. Iteratively reconfiguring the V_s or U_s can automatically find the right or left singular vectors, respectively, of an unknown transmission matrix $\mathcal{A}_s$. Gradually reconfiguring the row $i + 1$ (dashed boxes) successively maps the initial ith path-mode onto the corresponding supermode. The QAB and CCU processes are decomposed to a sequence of matrix operations on signal and idler photons: loss-balance Λ_{si}, phase compensation P_{si}, the identity I, and encoding-reversal R. (**C**) Simulated mutual information S of the transmitted multidimensional entangled states in four cases: without QER, with CCU, with CCU + R, and with the full QER. The dashed line indicates the ideal value for $|\Phi_4^+\rangle$: $S_{max} = 4$. (**D**) Experimental implementation of CCU. Normalized optical intensities for different modes were measured as the algorithm iteratively went on. Within 130 iterations, the CCU algorithm converged (when the variation of transmission was less than 5% for nearby iterations), and we interrupted the channel at step 130, which was online recovered again within 70 iterations. (Insets) Measured cross-talk matrices at steps 3, 30, 75, 129, and 198, and the element (i, j) denotes the cross-talk ratio from ith mode to jth mode.

is shown in Fig. 2A. We demonstrate in small scale the first multichip multidimensional entanglement networks, between three node chips and a server chip, linked by three FMFs (Fig. 2B, circuit diagram). To implement the entanglement network architecture, hybrid multiplexing devices are monolithically integrated. The networking chips allow wavelength-multiplexing and -demultiplexing of single photons; generation, manipulation, and measurement of path-encoded multidimensional entangled states; as well as coherent conversion to hybrid polarization mode–encoded states in FMFs. For scalable implementations, the chips require a broad bandwidth of spectral and modal modes, large fabrication tolerance, and high uniformity across the wafer. Broadband (a 3 dB bandwidth of 50 nm) biphoton correlations are generated by means of a spontaneous four-wave mixing (SFWM) process in silicon waveguides (Fig. 2C). By coherently pumping four sources, three pairs of four-dimensional path-encoded Bell state $|\Phi_4^+\rangle = 1/2\sum_{k=0}^{3}|k\rangle|k\rangle$ at paired wavelength $\{\lambda_{\pm3}, \lambda_{\pm2}, \lambda_{\pm1}\}$ are prepared, where $|k\rangle$ denotes a logical state in the kth path (*29*). Four arrays of coupled second-order microring resonators (MRRs; individually tunable) follow to demultiplex and remultiplex the correlated photons as $\{[\lambda_{-3}, \lambda_{-2}], [\lambda_{+1}, \lambda_{+2}], [\lambda_{-1}, \lambda_{+3}]\}$

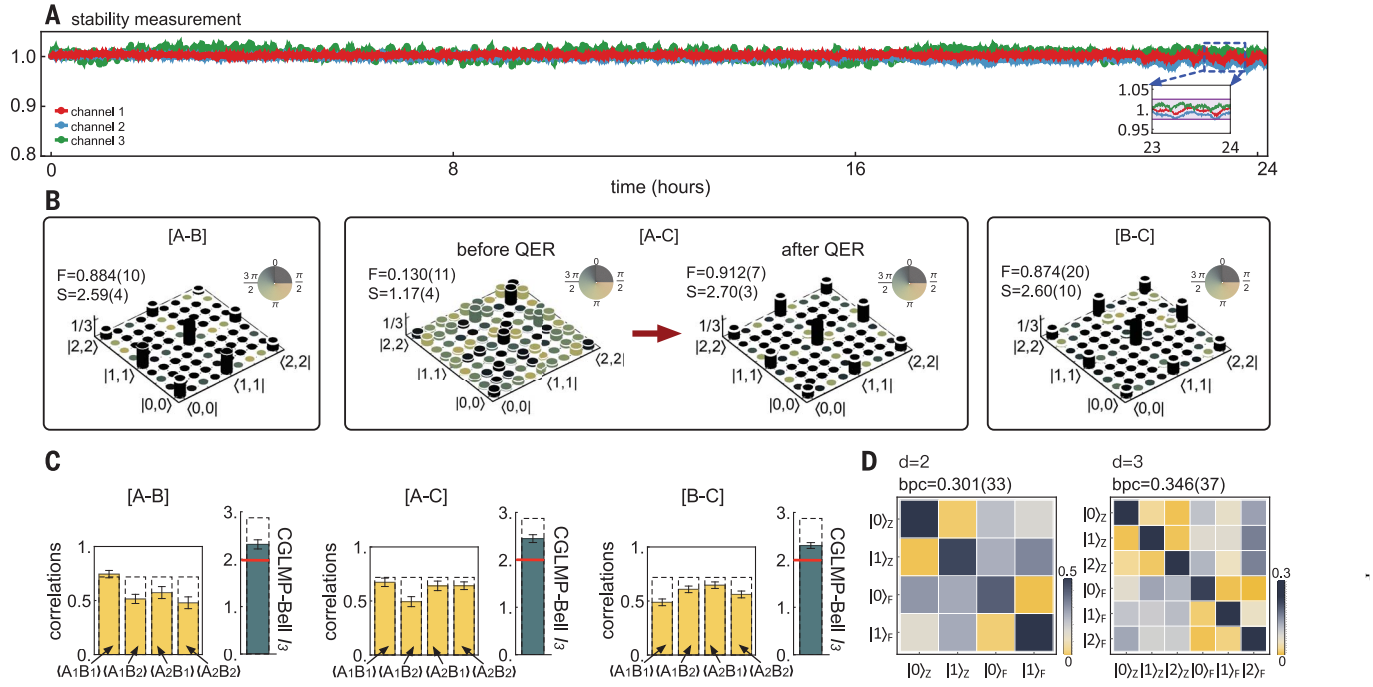


Fig. 4. Experimental validations of multidimensional entanglement distributions across multiple silicon chips. (**A**) Stability measurement. It is characterized by the normalized mode cross-talk $\left(O_{i,i}/\sum_{j=0}^{2}O_{i,j}\right)$, throughout the server chip, chip-fiber interface, to the FMF, and last, back to the node chip, where $O_{i,j}$ is the optical intensity collected at the *j*th port on the node chip by injecting light at the *i*th port on the server chip. Data were collected every 20 s. During measurement, the FMFs (about 1.5 m long each) were fixed to a table. (**B**) Experimentally reconstructed density matrices for the distributed multidimensional entangled states between node chips A and B, A and C, and B and C. The density matrix before entanglement retrieval between chips A and C is shown for comparison. Column heights indicate the absolute values |ρ|; colors indicate the phases Arg(ρ) mod 2π. Fidelities (*F*) and mutual information (*S*) are shown in the plots. The ideal value for $|\Phi_3^+\rangle$ is S_{max} = 3.17. (**C**) Experimental estimation of the generalized Bell-CGLMP parameters (I_3) and their quantum correlators $\langle A_a B_b\rangle$ between node chips. Dashed boxes indicate theoretical values. (**D**) Correlation matrices for multidimensional chip-fiber-chip QKD. *Z* indicates the computational basis, and *F* indicates the Fourier basis. The secure key rates in bits per coincidence are measured for the two- and three-dimensional entangled states. Color bars indicate the measured probabilities. Error bars in (C) and values in parentheses in (B) and (D) indicate ±1σ uncertainty, estimated from Poissonian photon statistics.

(Fig. 2B). Measured highly overlapped flat-top spectra of the MRRs arrays are shown in Fig. 2D (3 dB bandwidth of around 0.8 nm for each MRR); the wafer-scale test showed good fabrication tolerance of the design (*29*). We further characterized the photon pair sources by measuring reversed Hong-Ou-Mandel quantum interference. The measured visibilities [with a mean of 0.967(32) for all 18 interference fringes (*29*), where the number in parentheses is ±1σ uncertainty] confirm the high spectral indistinguishability of single photons after MRR demultiplexing.

The multipath-encoded states are inherently phase-stabilized on chip (*30*), whereas encoding *d*-dimensional states in a superposition of multiple eigenmodes in MMFs is preferred owing to relatively stable phases between internal modes. We devised a multidimensional chip-fiber interface that can coherently transform between on-chip path encoding and polarization-mode encoding in fiber. This interface consists of a mode multiplexer (*31*) and a two-dimensional multimode grating coupler (MMGC) (*32*). Designs are provided in (*29*), section 1. The coherent transformation between on-chip multipath-encoded states and polarization mode–encoded states in FMFs works as (*29*)

$$|0\rangle\leftrightarrow|TE_{1x}\rangle\leftrightarrow|LP_{11x}\rangle,\ |1\rangle\leftrightarrow|TE_{0x}\rangle\leftrightarrow|LP_{01x}\rangle$$

$$|2\rangle\leftrightarrow|TE_{0y}\rangle\leftrightarrow|LP_{01y}\rangle,\ |3\rangle\leftrightarrow|TE_{1y}\rangle\leftrightarrow|LP_{11y}\rangle$$

where $|TE_0\rangle$ and $|TE_1\rangle$ represent the logical states in the fundamental and first-order transverse electric (TE) modes of waveguides, $|LP_{01}\rangle$ and $|LP_{11}\rangle$ represent the logical states in the LP modes of FMFs, and *x* and *y* represent two orthogonal polarization states. On the node chip, the states are reversely transformed back to path encoding. Simulated and experimental LP-mode distributions at the output of FMF are shown in Fig. 2E [a full data set is provided in (*29*)]. The measured spectra from the single-mode input, to mode multiplexer, MMGC, and through FMF are shown in Fig. 2F. It shows a 3 dB bandwidth of about 15 nm and a mode-dependent loss of about 2.9 dB. Higher-order mode multiplexer—for example, up to 11 modes (*33*)—and high-performance multimode chip-fiber edge couplers (*34*) have been realized with multimode silicon photonics (*35*, *36*), and MMFs can easily support hundreds of LP modes—for example, 400 modes in a typical 50 μm-diameter MMF (*23*)—therefore promising substantial scalability of quantum network's dimension.

Multidimensional entanglement retrieving

Although MMFs show access to huge information capacity, lights that propagate in these complex media experience troublesome degradation that results in mode scrambling and entanglement degenerating (*19*, *20*). In our experiment and in most real-world systems, the transmission matrix of channels—including in this work the path-to-mode transformation and chip-fiber interface—is generally scrambled and nonunitary. This is caused by inevitable mode loss and cross-talk in MMFs because of fiber bending, twisting, or stretching, and mode mismatching at the interface. On the basis of Miller's scheme (*25*, *26*), we realized a QER method that can actively retrieve multidimensional entanglement from any complex scattering (nonunitary) process in the reciprocal system. By all-optically implementing the QER, the network can self-train its controllable

chips (node and server) to retrieve multidimensional entanglement without any need of reconstructing the matrix of channels and matrix processing [discussion is available in (*29*)]. The QER consists of two stages: a classical channel unscrambling (CCU) algorithm and a quantum amplitude balancing (QAB) strategy (Fig. 3A). The overall process is summarized in Fig. 3B. More details are provided in (*29*), section 4.

In the first stage, the CCU algorithm is carried out to unscramble the strongly mixed modes in chip-fiber-chip channels and reconstruct the orthogonal supermodes. This process can be regarded as the singular value decomposition (SVD) of an unknown nonunitary transmission matrix $\mathcal{A} = U\Sigma V^{\dagger}$, where Σ is a diagonal matrix with non-negative diagonal elements σ_k (singular values) arranged in descending order ($\sigma_0 \geq \sigma_1 \geq \ldots \geq \sigma_{d-1} \geq 0$), and V and U are a unitary matrix composed of the right-singular vector $\boldsymbol{v}_k$ and left-singular vector $\boldsymbol{u}_k$, respectively. The right and left singular vectors are dual invariant vectors of $\mathcal{A}$: $\mathcal{A}\boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$, $\mathcal{A}^T\overline{\boldsymbol{u}_k} = \sigma_k\overline{\boldsymbol{v}_k}$, where overlines denote complex conjugate. Hence, the superposition, in the form of right- and left-singular vectors, of initial encoding bases constitutes the orthogonal invariant input- and output-supermodes $|\boldsymbol{v}_k\rangle$ and $|\boldsymbol{u}_k\rangle$ of the channel, respectively, and singular values correspond to supermode-dependent losses.

We performed SVD on the transmission matrix and determined the invariant supermodes through bilateral power iterations (we label signal photon as "s" and the idler photon as "i" in the following), using two triangular linear optical circuits (V_s on the server chip and U_s on the node chips). The dominant supermode with singular value σ_0 could be reconstructed first. On the server chip, the light beam with unit intensity was injected into the first port $|0\rangle$ of V_s (initially set as identity), which can be expanded as the superposition of input-supermodes as $|\varphi_0\rangle = |0\rangle_{\text{server}} = \sum_{k=0}^{3} a_k|\boldsymbol{v}_k\rangle$, where a_k is the complex coefficient. The forward transmission of light through the FMF is equivalent to acting the transmission matrix $\mathcal{A}_s$ on the state $|\varphi_1\rangle = \mathcal{A}_s|\varphi_0\rangle = \sum_{k=0}^{3}\sigma_k a_k|\boldsymbol{u}_k\rangle$. Then, on the node chip, tunable beamsplitters and phase-shifters in row 1 of U_s were adjusted to minimize the output power $\{O_3, O_2, O_1\}$ successively, forcing all power into the $|0\rangle_{\text{node}}$ mode, $U_s^{(1)}|\varphi_1\rangle \propto |0\rangle_{\text{node}}$ (superscript is the iteration step). Then, the beam was backwardly sent into the first mode of the node chip. The reversed state is $|\overline{\varphi_s}\rangle = \mathcal{A}_s^T U_s^{(1)T}|0\rangle_{\text{node}} \propto \sum_{k=0}^{3}\sigma_k^2\overline{a_k}|\overline{\boldsymbol{v}_k}\rangle$ (the backward-transmission matrix is $\mathcal{A}_s^T$). We then adjusted V_s to get the $|0\rangle_{\text{server}}$. Physically repeating this all-optical process is equivalent to alternatively acting $\mathcal{A}_s$ and $\mathcal{A}_s^T$ on the vector. In the nth step, the added weight of kth supermode is σ_k^{2n}, which means that the nondominant components are exponentially compressed as iteration goes. When converging at fth iteration, it returns the dominant supermode $V_s^{(f)}|0\rangle_{\text{server}} \rightarrow |v_0\rangle$, $U_s^{(f)T}|0\rangle_{\text{node}} \rightarrow |\overline{\boldsymbol{u}_0}\rangle$ with the minimum loss. Last, repeating the bilateral power iterations for different input ports $|k\rangle$ allowed the mapping for all other supermodes with descending singular values (increasing loss).

Maximal entanglement, however, cannot be retrieved only through CCU. The signal and idler photons of the state $|\Phi_4^+\rangle$ were transmitted through two separate FMFs. Mapping from path-encoded states onto supermodes, the transmission matrices turn into a diagonal in the new basis: $\mathcal{A}_s \rightarrow \Sigma_s P_s$, $\mathcal{A}_i \rightarrow \Sigma_i P_i$, and the transmitted state becomes $|\Psi\rangle = (\Sigma_s P_s \otimes \Sigma_i P_i)|\Phi_4^+\rangle = 1/2\sum_{k=0}^{3}\sigma_k^{(s)}\sigma_k^{(i)}e^{i(\theta_k+\phi_k)}|k\rangle|k\rangle$, where Σ_s and Σ_i are matrices of singular values $\sigma_k^{(s)}$ and $\sigma_k^{(i)}$ in descending order, and P_s and P_i are redundant-phase matrices from the nonuniqueness of SVD: $P_s = \text{diag}\left[e^{i\theta_0}, e^{i\theta_1}, e^{i\theta_2}, e^{i\theta_3}\right]$ and $P_i = \text{diag}\left[e^{i\phi_0}, e^{i\phi_1}, e^{i\phi_2}, e^{i\phi_3}\right]$. The state $|\Psi\rangle$ remains partially entangled, and the amplitudes need to be balanced, for which QAB was used. QAB first reverses the encoding basis of one photon (such as the idler) on the server chip by operating a permutation matrix R that exchanges $|k\rangle$ and $|3-k\rangle$. That gives a new combination of two singular-value sets, and the state can be described as $|\Psi_1\rangle = 1/2\sum_k \sigma_k^{(s)}\sigma_{3-k}^{(i)}e^{i(\theta_k+\phi_{3-k})}|k\rangle|3-k\rangle$. The square of the factor $\sigma_k^{(s)}\sigma_{3-k}^{(i)}$ proportional to coincidence counts can be tailored to realize the maximally entangled state $|\Psi_2\rangle = 1/2\min\left\{\sigma_k^{(s)}\sigma_{3-k}^{(i)}\right\}\sum_k e^{i(\theta_k+\phi_{3-k})}|k\rangle|3-k\rangle$, by configuring pump distributions into the sources to alter the squeezing parameter or by slightly shifting the resonance of MRRs to adjust the spectral overlapping of photons (Fig. 3B, Λ_{si}). The value of $\min\left\{\sigma_k^{(s)}\sigma_{3-k}^{(i)}\right\}$ reaches maximum (minimum total loss) among all combinations of two singular-value sets for QAB (*29*). The phase factors can be simply compensated with phase shifters (Fig. 3B, P_{si}).

We adopted quantum mutual information to characterize the entanglement, defined as $S(A{:}\,B) = S_A + S_B - S_{AB}$, where $S_{A(B)} = -\text{Tr}[\rho_{A(B)}\text{Log}_2\rho_{A(B)}]$ is von Neumann entropy for the subsystem and S_{AB} is von Neumann entropy for the whole system. The non-negative S value decreases with less entanglement. We chose 100 pairs of random transmission matrices $\{\mathcal{A}_s, \mathcal{A}_i\}$ to validate the QER. Simulation results in different scenarios are shown in Fig. 3C. Implementing the full QER, the mutual information is recovered to the maximum value. The experimental results of unscrambling mixed modes in the multidimensional chip-fiber-chip channel by means of CCU are reported in Fig. 3D. The four supermodes were successively unscrambled from the one with the largest singular value (minimum loss) to the one with the smallest value (maximum loss). The recovered modes possessed much lower cross-talk. To show its online recovery ability and high robustness, which are key for practical quantum networking, we intentionally interrupted the FMF by bending the fiber and misaligning fiber-chip coupling (Fig. 3D, red triangle with exclamation point), and the algorithm surmounted it rapidly. Truncating lossy modes under the communication-capable threshold in larger-dimensional channels can further speed up the algorithm and reduce the total loss of the state. For this work, we truncated the lossy fourth mode to increase the count rates and prototyped the retrieved multidimensional entanglement network.

Entanglement distribution across multichip networks

We measured the stability of the whole system by monitoring the cross-talk matrix over a long time (Fig. 4A). Any occurrence of state rotations, whether on chips or in FMFs or their interface, could be monitored with this measurement. We observed fluctuation below ±2.5% within 24 hours, indicating a well-stabilized multidimensional chip-fiber-chip quantum system. No recalibration or realignment were needed during all the measurements in this work.

We then coherently distributed three pairs of three-dimensional entangled photons from the server chip to three node chips through three FMFs. We implemented the full QER method to retrieve entanglement. To verify the presence of retrieved entanglement, we implemented quantum tomographical measurements using compressed sensing techniques (*37*) to reconstruct the density matrices (ρ) and calculate the bipartite mutual information (S), after distributions across the multichip network. The results for reconstructed density matrices are provided in Fig. 4B. Quantum state fidelity is defined as $F = \langle\Phi_3^+|\rho|\Phi_3^+\rangle$, where $|\Phi_3^+\rangle$ is the ideal pure state, and the mutual information S is estimated from ρ. We obtained F = 0.884(10), S = 2.59(4) between node chips A and B, F = 0.912(7), S = 2.70(3) between node chips A and C, and F = 0.874(20), S = 2.60(10) between node chips B and C. As a comparison, in Fig. 4B, we show the density matrices before and after entanglement retrieval for node chips A and C, demonstrating the success of QER. The degradation is mainly due to the imperfect QER caused by some improvable defects of our devices (*29*). We tested a generalized Bell Collins-Gisin-Linden-Massar-Popescu (Bell-CGLMP) inequality (*38*) to certify the presence of multidimensional entanglement after distributions between all paired node chips in the network [(*29*), section 5]. The experimentally measured expected values of quantum correlators $\langle A_a B_b\rangle(\{a,b\} \in \{1,2\})$ and the measured Bell parameter I_3 are shown in Fig. 4C. We estimated I_3 of 2.31(9), 2.45(8), and 2.30(6) for the three pairs of the entangled

state, which confirm the experimental violations of the classical bound by 3.44σ, 5.63σ, and 5.00σ, respectively. The simultaneous violations of inequalities for all linked chips therefore confirm the success of multidimensional entanglement distributions in the multichip quantum network.

Additionally, our multichip system has the potential to provide higher information capacity and greater noise resistance for QKD. In Fig. 4D, we report the experimental correlation matrices for multidimensional entanglement–based QKD. The two chips in the network are measured in the mutually unbiased bases. We obtain higher bits per coincidence (bpc) of 0.346(37) for three-dimensional entangled states even with a larger quantum bit error rate (QBER; defined as the infidelity) of 10.3%, compared with 0.301(33) bpc and 4.6% QBER for two-dimensional entangled states, respectively. Details of multidimensional entanglement–based QKD are provided in (*29*), section 6.

Conclusions

We have demonstrated a multichip multidimensional quantum entanglement network with retrievability using silicon-photonic hybrid multiplexing technologies. The full spectrum of hybrid encoding and multiplexing devices have been monolithically integrated on chips, delivering one of most complicated integrated quantum photonic devices in terms of circuitry complexity and functionality. The scalabilities of networking architecture, integrated quantum devices, MMF channels, and entanglement retrieving techniques have been demonstrated and verified. In future real-world long-distance quantum networks, environment noises that result in mode cross-talk and phase drift in MMFs could be real-time all-optically corrected by using fast switching and detection of light in silicon (*29*). Our work points toward the practical implementations of large-scale, chip-based quantum entanglement networks for quantum information processing and communications (*1*, *2*). Technical advances in entanglement retrieving and integrated-photonics quantum information processing could provide a programmable way for operating quantum light in ultrahigh-dimensional fiber-based complex media, which could be amenable to fundamental and applied sciences (*19*, *20*).

REFERENCES AND NOTES

1. H. J. Kimble, *Nature* **453**, 1023–1030 (2008).
2. S. Wehner, D. Elkouss, R. Hanson, *Science* **362**, eaam9288 (2018).
3. B. Fröhlich *et al*., *Nature* **501**, 69–72 (2013).
4. S. Wang *et al*., *Opt. Express* **22**, 21739–21756 (2014).
5. Y.-A. Chen *et al*., *Nature* **589**, 214–219 (2021).
6. J. Wang, F. Sciarrino, A. Laing, M. G. Thompson, *Nat. Photonics* **14**, 273–284 (2020).
7. A. W. Elshaari, W. Pernice, K. Srinivasan, O. Benson, V. Zwiller, *Nat. Photonics* **14**, 285–298 (2020).
8. G. Moody *et al*., *J Phys Photonics* **4**, 012501 (2022).
9. P. Sibson *et al*., *Nat. Commun.* **8**, 13984 (2017).
10. G. Zhang *et al*., *Nat. Photonics* **13**, 839–842 (2019).
11. Y. Ding *et al*., *npj Quantum Inf.* **3**, 25 (2017).
12. K. Wei *et al*., *Phys. Rev. X* **10**, 031030 (2020).
13. T. K. Paraïso *et al*., *Nat. Photonics* **15**, 850–856 (2021).
14. J. Wang *et al*., *Optica* **3**, 407 (2016).
15. D. Llewellyn *et al*., *Nat. Phys.* **16**, 148–153 (2020).
16. X. Lu *et al*., *Nat. Phys.* **15**, 373–381 (2019).
17. X. Liu *et al*., *PhotoniX* **3**, 2 (2022).
18. W. Wen *et al*., *Phys. Rev. Appl.* **18**, 024059 (2022).
19. O. Lib, Y. Bromberg, *Nat. Phys.* **18**, 986–993 (2022).
20. H. Cao, A. P. Mosk, S. Rotter, *Nat. Phys.* **18**, 994–1007 (2022).
21. H. Defienne *et al*., *Opt. Lett.* **39**, 6090–6093 (2014).
22. Y. Zhou *et al*., *Nat. Commun.* **12**, 1866 (2021).
23. N. H. Valencia, S. Goel, W. McCutcheon, H. Defienne, M. Malik, *Nat. Phys.* **16**, 1112–1116 (2020).
24. B. Ndagano *et al*., *Nat. Phys.* **13**, 397–402 (2017).
25. D. A. B. Miller, *J. Lightwave Technol.* **31**, 3987–3994 (2013).
26. A. Annoni *et al*., *Light Sci. Appl.* **6**, e17110 (2017).
27. S. Wengerowsky, S. K. Joshi, F. Steinlechner, H. Hübel, R. Ursin, *Nature* **564**, 225–228 (2018).
28. S. K. Joshi *et al*., *Sci. Adv.* **6**, eaba0959 (2020).
29. Materials and methods are available as supplementary materials.
30. J. Wang *et al*., *Science* **360**, 285–291 (2018).
31. D. Dai *et al*., *Laser Photonics Rev.* **12**, 1700109 (2018).
32. Y. Tong, W. Zhou, X. Wu, H. K. Tsang, *IEEE J. Quantum Electron.* **56**, 1–7 (2020).
33. Y. He *et al*., *J. Lightwave Technol.* **36**, 5746–5753 (2018).
34. Y. Lai *et al*., *Opt. Lett.* **42**, 3702–3705 (2017).
35. I. Cristiani *et al*., *J. Opt.* **24**, 083001 (2022).
36. C. Li, D. Liu, D. Dai, *Nanophotonics* **8**, 227–247 (2019).
37. C. A. Riofrío *et al*., *Nat. Commun.* **8**, 15305 (2017).
38. D. Collins, N. Gisin, N. Linden, S. Massar, S. Popescu, *Phys. Rev. Lett.* **88**, 040404 (2002).

ACKNOWLEDGMENTS

Funding: We acknowledge support from the Innovation Program for Quantum Science and Technology (2021ZD0301500), The Natural Science Foundation of China (62235001, 61975001, 62125503, 91950205, 61961146003, and 62205292), National Science Fund for Distinguished Young Scholars (61725503), the National Key R&D Program of China (2019YFA0308702), the Fundamental Research Funds for the Central Universities, the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2021R01001), Beijing Natural Science Foundation (Z190005 and Z220008), Key R&D Program of Guangdong Province (2018B030329001), and Hong Kong RGC GRF grant (14203620). **Author contributions:** J.W. and D.D. conceived the project. Y.Z., C.Z., and D.L. contributed equally to this work. Y.Z., C.Z., D.L., X.C., Z.F., Y.T., X.Z., H.K.T., D.D., and J.W. designed the chips. Y.Z., C.Z., D.L., J.M., T.D., J.H., and J.B. built the setup and carried out the experiment. J.M., Y.Y., B.T., and Z.L. fabricated the device. Y.Z., C.Z., D.L., J.H., Z.F., Y.T., and X.Z. performed the theoretical analysis. Y.L., Q.G., H.K.T., D.D., and J.W. managed the project. Y.Z., C.Z., D.L., D.D., and J.W. wrote the manuscript. **Competing interests:** The multimode waveguide grating coupler is protected by US patent "US-11428869-B2." **Data and materials availability:** All data are available in the manuscript or in the supplementary materials. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg9210
Materials and Methods
Figs. S1 to S10
References (*39*–*57*)

Submitted 31 January 2023; accepted 5 June 2023
10.1126/science.adg9210

SUPERCONDUCTIVITY

Correlating the charge-transfer gap to the maximum transition temperature in $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$

Zechao Wang[1,2]†, Changwei Zou[3]†, Chengtian Lin[4], Xiangyu Luo[5], Hongtao Yan[5], Chaohui Yin[5], Yong Xu[3,6,7], Xingjiang Zhou[5], Yayu Wang[3,6]*, Jing Zhu[1,2]*

As the number of CuO_2 layers, n, in each unit cell of a cuprate family increases, the maximum transition temperature ($T_{c,max}$) exhibits a universal bell-shaped curve with a peak at $n = 3$. The microscopic mechanism of this trend remains elusive. In this study, we used advanced electron microscopy to image the atomic structure of cuprates in the $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$ family with $1 \le n \le 9$; the evolution of the charge-transfer gap size (Δ) with n can be measured simultaneously. We determined that the n dependence of Δ follows an inverted bell-shaped curve with the minimum Δ value at $n = 3$. The correlation between Δ, n, and $T_{c,max}$ may clarify the origin of superconductivity in cuprates.

Identifying universal trends of the superconducting transition temperature (T_c) and its correlations to other physical parameters may provide crucial clues for elucidating the origin of superconductivity (*1*). For example, the isotope effect $T_c \propto 1/\sqrt{M}$, where M is the isotopic mass, has inspired the phonon-mediated pairing picture in establishing the microscopic theory of conventional superconductivity in metals and alloys (*2*, *3*). Since the discovery of high-T_c superconductivity in copper oxide materials (referred to as cuprates), there have been considerable efforts in understanding what controls T_c. There are two well-established trends about T_c in cuprates: One is the dome-shaped doping dependence of T_c for a specific cuprate compound, where the maximum T_c ($T_{c,max}$) is located at hole concentration $p \sim 0.16$ (*1*). The other is the variation of $T_{c,max}$ with the number of CuO_2 planes (n) per unit cell in a homologous series, which reaches the highest when $n = 3$ (*4*). Understanding the physical origin behind these empirical rules could lead to the eventual solution for the pairing-mechanism problem in cuprates (*1*, *4*).

Whereas the second trend remains universal for many cuprate families (Fig. 1A), the first is apparently violated in the trilayer cuprate $Bi_2Sr_2Ca_2Cu_3O_{10+\delta}$ (Bi-2223), which has the highest $T_{c,max}$ of ~115 K in the $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$ family (Bi-family) compounds. Recent scanning tunneling microscopy (STM) studies (*5*) have shown that the superconducting gap of Bi-2223 decreases monotonously with p in the overdoped regime. Unexpectedly, the T_c in this doping range exhibits a plateau rather than a continuous drop, most likely owing to the existence of two types of inequivalent CuO_2 planes (*6*, *7*). By contrast, the Bi-2223 system provides an ideal platform for investigating the evolution of $T_{c,max}$ with n and the physical parameters that control $T_{c,max}$. A promising starting point is the Mott-insulating "parent" state ($p \sim 0$) of cuprates, where the electronic structure can be described by a single parameter referred to as the charge-transfer gap (CTG) Δ, the size of which varies from 1 to 3 eV (*8*). A previous STM work (*9*) in $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$ and $Ca_{n+1}Cu_nO_{n+2}Cl_2$ with $n = 1$ and 2 suggests that $T_{c,max}$ is anticorrelated with the Δ, together with theoretical calculations (*10*–*12*). However, a measurement covering the whole homologous cuprate series has not yet been achieved because of two main obstacles. First, it is difficult to synthesize high-quality single crystals for $n \ge 4$ in any cuprate family, and it is even more difficult to reach the $p \sim 0$ limit. Second, most of the spectral information about cuprates cannot distinguish the distinct CuO_2 planes for $n \ge 3$.

The development of advanced electron microscopy enables one to achieve through-depth (defined in this paper as along the crystallographic c axis) imaging and carry out high-spatial (subangstrom)–resolution electron energy–loss spectrometry (EELS) experiments. Thus, it allows a local probe of the intrinsic structural and electronic properties in confined unit cells (*13*). Here, we used scanning transmission electron microscopy (STEM) and EELS techniques to directly image the layer-by-layer lattice configuration and electronic structure of the Bi-family cuprates, covering an unprecedented layer structure range with $1 \le n \le 9$.

Imaging the atomic structure of the Bi-family compounds with $1 \le n \le 9$.

The evolution of $T_{c,max}$ with n for the Bi-family cuprates is indicated with a bold orange line in Fig. 1A, which displays the universal bell-shaped trend (*14*–*19*). Owing to the high spatial resolution of the STEM technique, the layered atomic structure can be clearly visualized, as shown in Fig. 1B for an optimally doped Bi-2223 sample. Each Cu atom in the inner CuO_2 plane (IP) forms an in-plane CuO_4 plaquette, whereas in the outer CuO_2 planes (OP), the configuration is a CuO_5 pyramid with one apical oxygen (Fig. 1B, right). Some previous studies (*20*, *21*) have suggested that the different environments of IP and OP may have a profound influence on their superconducting properties, but their layer-resolved electronic properties remain to be unveiled.

Figure 1, C to K, displays a series of high-quality cross-sectional STEM images, in which $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$ with $1 \le n \le 9$ are all identified. We obtained the data for $n \ge 3$ by carefully searching within a dozen Bi-2223 samples, which include very scarce regions that have four or more CuO_2 planes per unit cell. The details about the search and discovery of the $n > 3$ phases are shown in fig. S5. Because of the quasi-two-dimensional nature of Bi-family cuprates, it has been shown recently that all the essential electronic properties are contained within a unit cell, BiO/SrO charge reservoir layers plus the CuO_2 layers (*22*). The methodology used here opens a gate for studying multilayer cuprates.

Measuring the charge-transfer gap of the Bi-family compounds.

The electronic band structure of cuprates is well characterized by the in-plane Cu $3d_{x^2-y^2}$ and doped O $2p_x/2p_y$ orbitals (*23*). Within the Zaanen-Sawatzky-Allen scheme (*24*), the undoped cuprate is a charge-transfer–type Mott insulator, in which the lowest-energy excitation is from the O charge-transfer band (CTB) to the unoccupied upper Hubbard band (UHB) of the Cu sites, which is schematically illustrated in Fig. 2A. Figure 2B shows the STM spectra in three insulating Bi-family cuprates with $n = 1$, 2, and 3, in which the Δ value between the edges of CTB and UHB can be clearly extracted. The results here expand on those of previous studies (*9*) to $n = 3$ and reveal the same trend of a monotonic decrease of Δ from 1.5 eV in Bi-2201 to 1.0 eV in Bi-2212 and 0.7 eV in Bi-2223.

Upon hole doping, an additional feature emerges from the CTB that is known as the Zhang-Rice singlet (ZRS) band (Fig. 2C). This feature can be attributed to the hybridized state of a doped hole involving one Cu-3d and the four nearest O-2p orbitals (*25*). In this circumstance, the main features in the STM spectra are the spectral weight transfer from high energy to low energy and the gradual formation of the pseudogap (*26*); the Δ value becomes less well defined. Nevertheless, the energy difference between the centers of the ZRS

[1]National Center for Electron Microscopy in Beijing, School of Materials Science and Engineering, Key Laboratory of Advanced Materials (MOE), The State Key Laboratory of New Ceramics and Fine Processing, Tsinghua University, Beijing, P.R. China. [2]Ji Hua Laboratory, Foshan, Guangdong, P.R. China. [3]State Key Laboratory of Low-Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing, P.R. China. [4]Max Planck Institute for Solid State Research, Stuttgart, Germany. [5]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, P.R. China. [6]New Cornerstone Science Laboratory, Frontier Science Center for Quantum Information, Beijing, P.R. China. [7]RIKEN Center for Emergent Matter Science (CEMS), Wako, Japan.
†These authors contributed equally to this work.
*Corresponding author. Email: jzhu@mail.tsinghua.edu.cn (J.Z.); yayuwang@mail.tsinghua.edu.cn. (Y.W.)

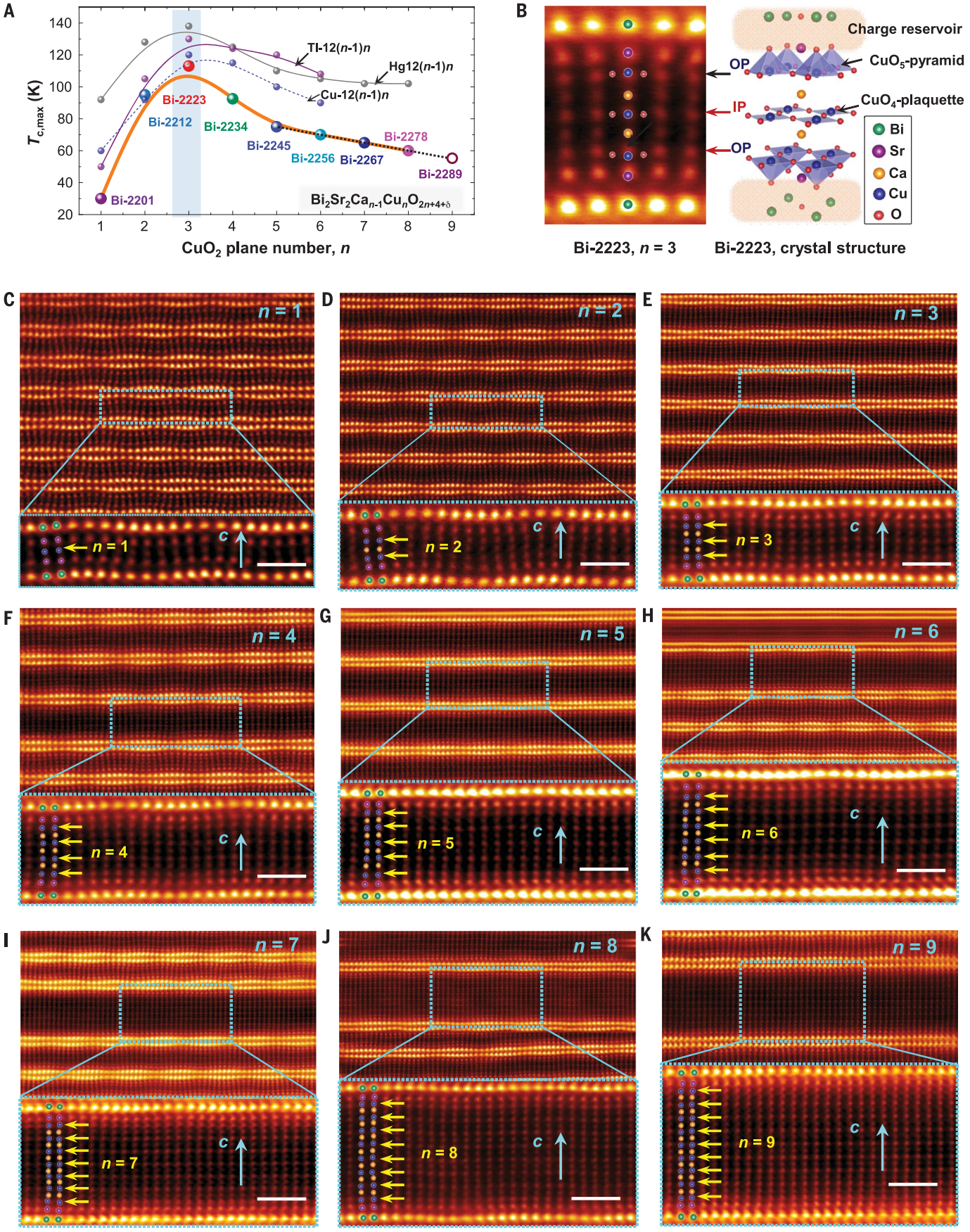

A
$T_{c,max}$ (K)
CuO$_2$ plane number, n
Tl-12(n-1)n
Hg12(n-1)n
Cu-12(n-1)n
Bi-2223
Bi-2212
Bi-2234
Bi-2245
Bi-2256
Bi-2267
Bi-2278
Bi-2289
Bi-2201
Bi$_2$Sr$_2$Ca$_{n-1}$Cu$_n$O$_{2n+4+\delta}$
B
Charge reservoir
OP
CuO$_5$-pyramid
IP
CuO$_4$-plaquette
OP
Bi
Sr
Ca
Cu
O
Bi-2223, n = 3
Bi-2223, crystal structure
C
n = 1
D
n = 2
E
n = 3
F
n = 4
G
n = 5
H
n = 6
I
n = 7
J
n = 8
K
n = 9
c

Figure 1

Fig. 1. The $T_{c,max}$ and atomically resolved crystal structure of $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$ cuprates with $1 \le n \le 9$. (**A**) Relationship between $T_{c,max}$ and n for the homologous series of n-layered cuprates (*14*–*19*) including Tl-12(n-1)n, Hg-12(n-1)n, Cu-12(n-1)n, and Bi-22(n-1)n. The bold orange line indicates the $T_{c,max}$ as a function of n in the Bi-family. The n = 9 phase has never been discovered, so the $T_{c,max}$ value was obtained with linear extrapolation (as indicated with the black dotted line) and represented by a hollow circle. (**B**) A zoomed-in area of the n = 3 sample (Bi-2223) with the schematic crystal structure shown at right. OP, outer CuO_2 plane; IP, inner CuO_2 plane. (**C** to **K**) Cross-sectional layered structure in the Bi-family cuprates with $1 \le n \le 9$, where the c axis direction is along the vertical blue arrow. The yellow arrows indicate the positions of CuO_2 planes in the crystal (fig. S4). Scale bar, 5 Å. All the data were taken at T = 300 K.

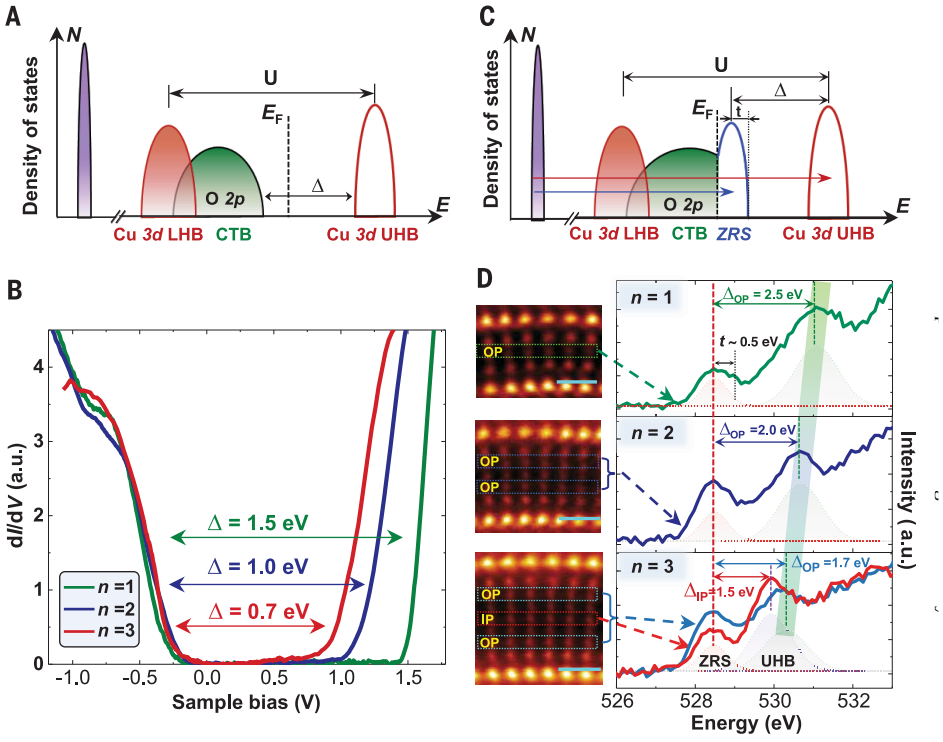


Fig. 2. Probing Δ with STM and STEM-EELS techniques. (**A**) Schematic band structure of the undoped cuprates. LHB, lower Hubbard band; CTB, charge transfer band; UHB, upper Hubbard band. (**B**) The Mott-insulator type d*I*/d*V* (differential conductance) spectra of undoped Bi-2201 (n = 1), Bi-2212 (n = 2), and Bi-2223 (n = 3) samples taken with STM. The Δ here characterizes the distance between the edges of CTB and UHB, as shown in (A). (**C**) Schematic band structure of hole-doped cuprates. ZRS, Zhang-Rice singlet. The red and blue arrows indicate the EELS excitation process from core levels to the unoccupied UHB and ZRS, respectively (the principle is shown in fig. S6). (**D**) STEM-EELS of the O *K* edge from n = 1 to 3. The Δ here characterizes the distance between the centers of ZRS and UHB, as shown in (C). The Gaussian functions were used to simulate the ZRS (red dotted lines) and the UHB (purple, IP; green, OP). (Left) The STEM images show the real-space areas where the EELS are taken. All the STEM-EELS data were taken at T = 300 K. a.u., arbitrary units.

and UHB is still a valid parameter that characterizes the charge-transfer energy (*27*), which varies little with doping (*28*, *29*). The state-of-the-art STEM-EELS is an ideal technique for measuring such peak-to-peak CTG, where the spectral peaks shown in Fig. 2D indicate the large probability of exciting core-level electrons to the unoccupied ZRS and UHB in Fig. 2C (*27*), in analogy to the x-ray absorption spectroscopy (XAS) at the O *K* edge (excitation energy of an O 1*s* core electron to an empty state). In particular, the ZRS peak centered around 528.5 eV (Fig. 2D, red dashed line) corresponds well with the measurements of the bulk-sensitive XAS technique (*27*). The ZRS peak is robust in cuprates against doping, temperature, and materials.

Similar to that of STM, the Δ obtained with EELS is mainly controlled by the unoccupied UHB, which displays a red-shift trend from n = 1 to 3, as revealed in Fig. 2D. We extracted the Δ size from a reliable fitting procedure as described in (*30*), which renders Δ = 2.5 eV for Bi-2201 and Δ = 2.0 eV for Bi-2212, respectively. Owing to the high atomic-plane resolution, we distinguished the CTG corresponding to the OP (Δ_{OP} = 1.7 eV) and IP (Δ_{IP} = 1.5 eV) in trilayer Bi-2223. The spectral weight ratio between the ZRS and UHB is smaller in the IP, which is suggestive of the lower hole concentration (*27*).

Correlating observables with the number of CuO_2 layers.

To investigate how the Δ size evolves with the number of CuO_2 layers per unit cell, we have further determined the STEM-EELS of the O *K* edge with $1 \le n \le 9$ (Fig. 3, A and B). The definition and description of different CuO_2 planes where we took the STEM-EELS measurements are shown in fig. S7. We repeated the measurements 10 times on different areas with the same n to improve statistical robustness (full dataset is provided in figs. S8 to S17) (*30*). For clarity, we only show the averaged spectra for the OP layer; those for the IP layers are shown in fig. S19. The overall trend is shown in Fig. 3, C and D, which clearly demonstrate that the n dependence of Δ is represented by an inverted bell-shaped curve with the minimum Δ = 1.8 eV at n = 3. There is an apparent anticorrelation between the n dependence of Δ and $T_{c,max}$. For all the phases in which $n \ge 3$, the Δ value of the IP is smaller than the Δ value of OP, but exhibits the same dependence on n (Fig. 3D).

In the scenario of single-band Hubbard model, we may define an effective superexchange energy $J_{eff} \sim 4t^2_{eff}/\Delta$, where t_{eff} is the effective hopping term that is proportional to t^2_{pd}/Δ, and t_{pd} is the hopping integral between neighboring O 2*p* and Cu 3*d* orbitals (*8*). Because the J_{eff} between local Cu moments has been proposed to be responsible for the spin singlet pairing (*31*), $J_{eff} \sim 1/\Delta^3$ thus correlates the CTG and the Cooper pairing strength. For n = 1 and 2, only the OP is present, and the increase of J_{OP} is consistent with that derived through STM (*9*) and inelastic photon scattering (*32*). For the whole range with $1 \le n \le 9$, Fig. 3E shows that the n dependence of J_{eff} for the OP layer exhibits a bell shape highly analogous to that of $T_{c,max}$. It reaches a peak at n = 3 and decreases for $4 \le n \le 9$. The close correlation between J_{eff} and $T_{c,max}$ indicates that the Δ, and thus the underlying strong correlation effects, determine $T_{c,max}$, at least in the Bi-family cuprates with $1 \le n \le 9$.

The combined STEM-EELS and STM results here touch on the central issue regarding the bell-shaped evolution of $T_{c,max}$ with n, which is universal in all cuprate families. From the theoretical point of view, a previous study (*4*)

has proposed that the enigmatic balance between Josephson tunneling and competing electronic orders is the underlying physical mechanism. Our results reveal that it is the CTG, which displays an inverted bell-shaped trend compared with the $T_{c,max}$ that plays a deterministic role in this empirical rule. This conclusion is supported by the consistency of Δ characterized by energy scales between band centers (EELS and XAS) and edges (STM). We summarize the CTG sizes taken with different experimental probes in Fig. 4. The difference between STM and EELS shows a rigid shift of ~1 eV, which is caused by the finite bandwidths and provides useful information about the

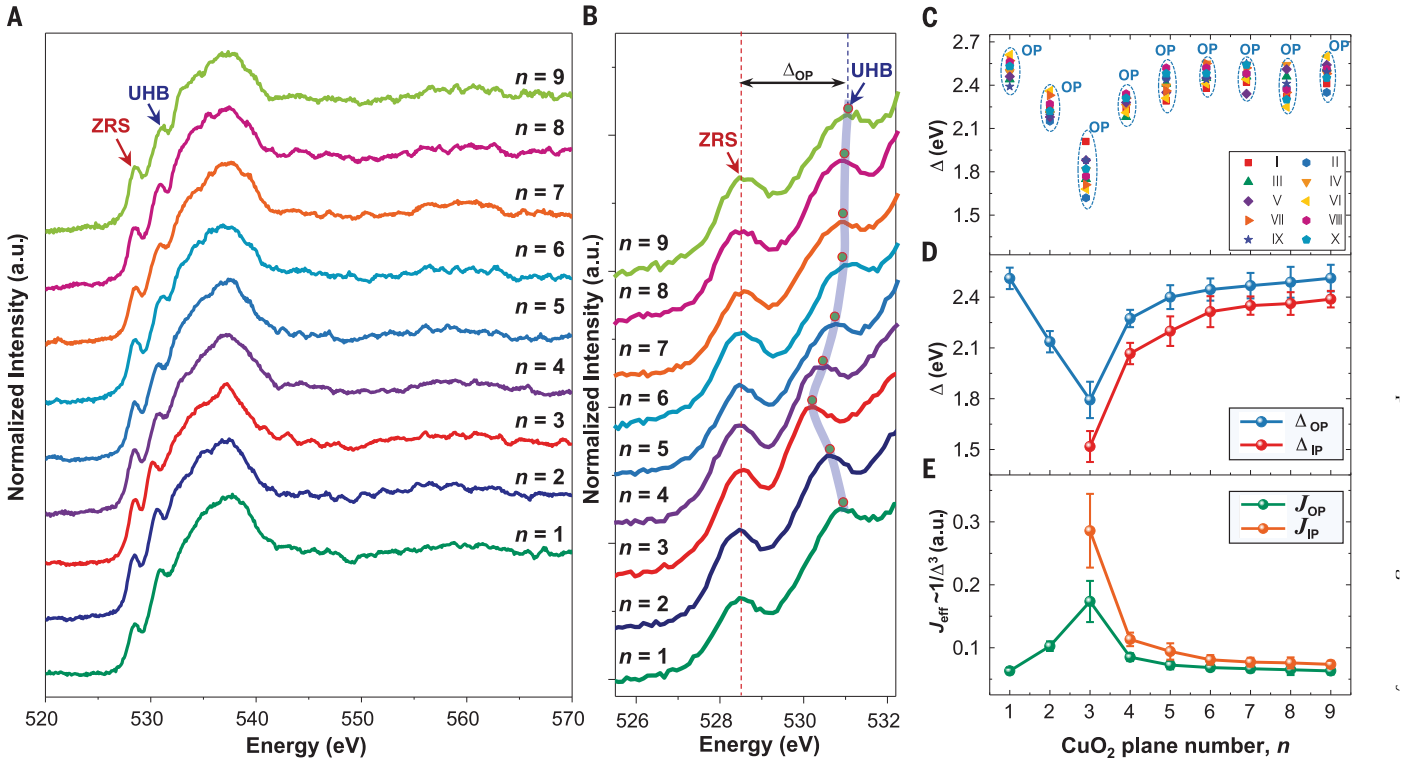


Fig. 3. The evolution of Δ with *n* in Bi-family cuprates. (**A**) STEM-EELS with a higher energy range of the O *K* edge with $1 \le n \le 9$. For each curve, we averaged the data measured on the OP layer of 10 different areas with the same *n* (full datasets are provided in figs. S8 to S17). (**B**) Zoomed-in STEM-EELS of (A). The Gaussian functions were used to determine the position of ZRS and UHB (fig. S18). (**C** and **D**) The statistical results of Δ for the OP and IP with $1 \le n \le 9$, which was determined with 10 measurements taken under the same experimental conditions. All the data were taken at T = 300 K. Each dot and error bar in (D) are the calculated average and standard deviation of the result in (C) for the same *n*. (**E**) The effective superexchange $J_{eff} \sim 1/\Delta^3$ as a function of *n*, assuming the same t_{eff} for the OP and IP of different *n* in a homologous series. a.u., arbitrary units.

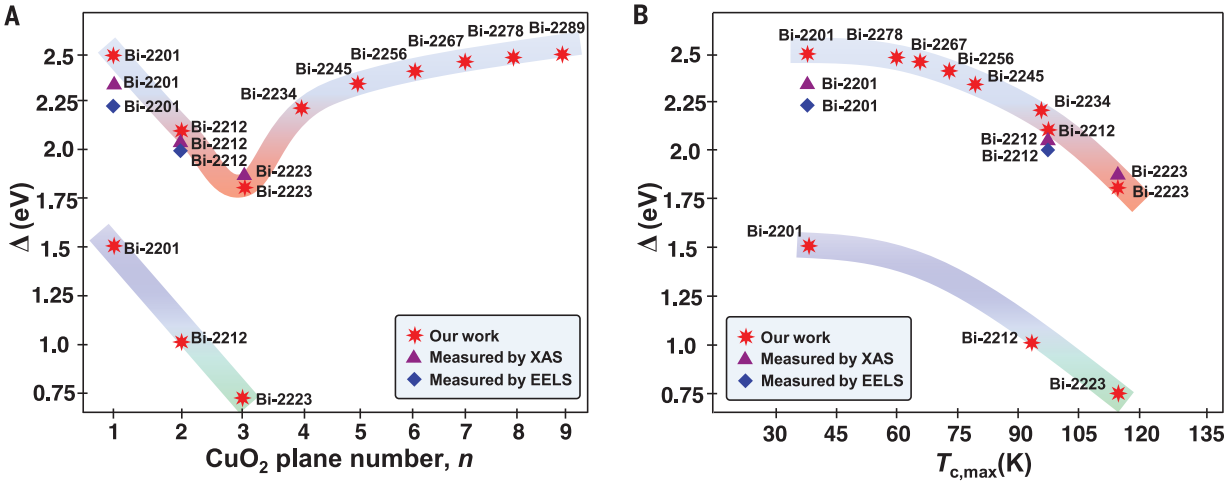


Fig. 4. The correlation between Δ, *n*, and $T_{c,max}$. (**A** and **B**) Summarized plots of (A) Δ versus *n* and (B) Δ versus $T_{c,max}$ for the Bi-family cuprates. Our STM and EELS (on the OP) data, as well as previous XAS data, are shown together. The STM results are generally ~1 eV smaller than the EELS and XAS values, owing to the finite bandwidth or hopping integral *t*. The wide gray-red and blue-green strips are guides to the eye. See table S1 for more details.

effective hopping integral in the Bi-family cuprates.

Discussion and outlook

Our study shows that the CuO_2 planes with different apical environments have very different Δ, which is seen not only in the evolution with n but also in that the IP always has a smaller Δ size than the OP of the same compound. The J_{eff} of IP is thus larger than that of OP, as shown in Fig. 3E, which is consistent with previous studies of the pairing strength (*21*, *33*). Therefore, the ultimate reason why Δ, and hence $T_{c,max}$, varies substantially for different n in a homologous series is most likely related to the orbitals out of the CuO_2 plane, indicating an interplay between the underlying electronic structure and orbital parameter symmetry in cuprates (*4*, *34*, *35*).

REFERENCE AND NOTES

1. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
2. C. A. Reynolds, B. Serin, W. H. Wright, L. B. Nesbitt, *Phys. Rev.* **78**, 487 (1950).
3. E. Maxwell, *Phys. Rev.* **78**, 477 (1950).
4. S. Chakravarty, H.-Y. Kee, K. Völker, *Nature* **428**, 53–55 (2004).
5. C. W. Zou *et al*., *Nat. Phys.* **18**, 551–557 (2022).
6. T. Fujii, T. Watanabe, A. Matsuda, *Physica C Supercond* **357–360**, 173–176 (2001).
7. Z. Hao *et al*., *Phys. Rev. Lett.* **125**, 237005 (2020).
8. P. A. Lee, N. Nagaosa, X. G. Wen, *Rev. Mod. Phys.* **78**, 17–85 (2006).
9. W. Ruan *et al*., *Sci. Bull.* **61**, 1826–1832 (2016).
10. C. Weber, C. Yee, K. Haule, G. Kotliar, *Europhys. Lett.* **100**, 37001 (2012).
11. N. Kowalski, S. S. Dash, P. Sémon, D. Sénéchal, A.-M. Tremblay, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106476118 (2021).
12. Z. H. Cui *et al*., *Phys. Rev. Res.* **2**, 043259 (2020).
13. D. A. Muller *et al*., *Science* **319**, 1073–1076 (2008).
14. H. Mukuda *et al*., *J. Phys. Soc. Jpn.* **81**, 011008 (2012).
15. C. W. Chu, L. Z. Deng, B. Lv, *Physica C* **514**, 290–313 (2015).
16. *Bismuth-Based High-Temperature Superconductors*, H. Maeda, Ed. (CRC Press, 1996).
17. H. Wang *et al*., *Appl. Phys. Lett.* **57**, 710–711 (1990).
18. H. Narita, T. Hatano, K. Nakamura, *J. Appl. Phys.* **72**, 5778–5785 (1992).
19. I. Bozovic, J. N. Eckstein, G. F. Virshup, *Physica C* **235**, 178–181 (1994).
20. C. Zou *et al*., *Phys. Rev. Lett.* **124**, 047003 (2020).
21. S. Kunisada *et al*., *Science* **369**, 833–838 (2020).
22. Y. Yu *et al*., *Nature* **575**, 156–163 (2019).
23. V. J. Emery, *Phys. Rev. Lett.* **58**, 2794–2797 (1987).
24. J. Zaanen, G. A. Sawatzky, J. W. Allen, *Phys. Rev. Lett.* **55**, 418–421 (1985).
25. F. C. Zhang, T. M. Rice, *Phys. Rev. B Condens. Matter* **37**, 3759–3761 (1988).
26. P. Cai *et al*., *Nat. Phys.* **12**, 1047–1051 (2016).
27. C. T. Chen *et al*., *Phys. Rev. Lett.* **66**, 104–107 (1991).
28. Y. Zhong *et al*., *Phys. Rev. Lett.* **125**, 077002 (2020).
29. C. Hu *et al*., *Nat. Commun.* **12**, 1356 (2021).
30. Materials and methods and additional data are available as supplementary materials.
31. P. W. Anderson, *Science* **235**, 1196–1198 (1987).
32. L. Wang *et al*., *Nat. Commun.* **13**, 3163 (2022).
33. H. Kotegawa *et al*., *Phys. Rev. B Condens. Matter* **64**, 064515 (2001).
34. E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen, O. K. Andersen, *Phys. Rev. Lett.* **87**, 047003 (2001).
35. T. Xiang, J. M. Wheatley, *Phys. Rev. Lett.* **77**, 4632–4635 (1996).
36. Z. C. Wang *et al*., Source Data for: Correlating the charge transfer gap to the maximum transition temperature in $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4+\delta}$, version 1, Zenodo (2023); https://doi.org/10.5281/zenodo.7608360.

ACKNOWLEDGMENTS

We gratefully acknowledge Q. K. Xue for helpful discussions and assistance. **Funding:** This work was financially supported by the Chinese National Natural Science Foundation (Basic Science Center Project of NSFC, grant no. 52388201), the Basic and Applied Basic Research Major Programme of Guangdong Province, China (grant no. 2021B0301030003), and Jihua Laboratory (project no. X210141TL210). Y.W. is supported by the Innovation Program for Quantum Science and Technology (grant no. 2021ZD0302502) and the New Cornerstone Science Foundation through the New Cornerstone Investigator Program and the XPLORER PRIZE. X.Z. is supported by the Chinese National Natural Science Foundation of China (grant no. 11888101), the National Key Research and Development Program (grant no. 2021YFA1401800), and the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (grant no. XDB25000000). This work used the resources of the National Center for Electron Microscopy in Beijing. **Author contributions:** J.Z. and Z.W. initiated the idea and designed the studies. Z.W. performed the electron microscopy experiments and data analysis with the help of J.Z. C.L. grew the crystals of Bi-2223. C.Z., X.L., H.Y., C.Y. and X.Z. processed the samples. C.Z. and Y.W. performed the STM experiment. All authors contributed to the scientific discussions. Y.X. offered some theoretical guidance on the DFT calculation. Z.W., C.Z., Y.W., and J.Z. wrote the paper with contributions from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are deposited in a public database of Zenodo (*36*).

License information: Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add3672
Materials and Methods
Figs. S1 to S20
Table S1
References (*37*–*47*)

Submitted 7 June 2022; accepted 6 June 2023
10.1126/science.add3672

HEART DISEASE

Recruited macrophages elicit atrial fibrillation

Maarten Hulsmans[1,2]*†, Maximilian J. Schloss[1,2]†, I-Hsiu Lee[1,2]†, Aneesh Bapat[3], Yoshiko Iwamoto[1,2], Claudio Vinegoni[1,2], Alexandre Paccalet[1,2], Masahiro Yamazoe[1,2], Jana Grune[1,2], Steffen Pabel[1,2,4], Noor Momin[1,2], Hana Seung[1,2], Nina Kumowski[1,2], Fadi E. Pulous[1,2], Daniel Keller[5], Constanze Bening[5], Ursula Green[6], Jochen K. Lennerz[6], Richard N. Mitchell[7], Andrew Lewis[8,9], Barbara Casadei[8,9], Oriol Iborra-Egea[10], Antoni Bayes-Genis[10], Samuel Sossalla[4,11], Chin Siang Ong[12,13], Richard N. Pierson[12], Jon C. Aster[7], David Rohde[1,2], Gregory R. Wojtkiewicz[1,2], Ralph Weissleder[1,2], Filip K. Swirski[14,15], George Tellides[13], George Tolis Jr.[16], Serguei Melnitchouk[12], David J. Milan[17], Patrick T. Ellinor[3,18], Kamila Naxerova[1,2]*‡, Matthias Nahrendorf[1,2,3,19]*‡

Atrial fibrillation disrupts contraction of the atria, leading to stroke and heart failure. We deciphered how immune and stromal cells contribute to atrial fibrillation. Single-cell transcriptomes from human atria documented inflammatory monocyte and *SPP1*$^{+}$ macrophage expansion in atrial fibrillation. Combining hypertension, obesity, and mitral valve regurgitation (HOMER) in mice elicited enlarged, fibrosed, and fibrillation-prone atria. Single-cell transcriptomes from HOMER mouse atria recapitulated cell composition and transcriptome changes observed in patients. Inhibiting monocyte migration reduced arrhythmia in *Ccr2*$^{-/-}$ HOMER mice. Cell-cell interaction analysis identified SPP1 as a pleiotropic signal that promotes atrial fibrillation through cross-talk with local immune and stromal cells. Deleting *Spp1* reduced atrial fibrillation in HOMER mice. These results identify SPP1^{+} macrophages as targets for immunotherapy in atrial fibrillation.

In healthy atria, excitations arising in the sinus node coordinate atrial contractions that help fill the ventricles in diastole. Cardiomyocytes constitute 30% of atrial cells (*1*). The remaining 70% of atrial cells are fibroblasts, endothelial cells, and resident immune cells. Together, these provide cardiomyocytes with housekeeping functions, including matrix scaffold, blood supply, waste removal, and immunosurveillance (*2*, *3*). Atrial fibrillation (AFib) is the most common arrhythmia (*4*). During AFib, the lack of coordinated atrial excitation and contraction slows blood flow into the ventricles. This sluggishness can diminish cardiac performance and give rise to atrial blood clots that cause thromboembolic stroke.

Inflammation may contribute to AFib through electrical remodeling (*5*–*7*), during which altered ion handling perturbs the action potential (*8*). Fibrosis, another consequence of inflammation, induces heterogeneous atrial depolarization, which provides an opportunity for electrical conduction reentry and subsequent AFib (*8*, *9*). To examine how immune and stromal cells contribute to AFib, we performed single-cell RNA sequencing (scRNA-seq) of freshly isolated normal and diseased human atrial tissue.

Macrophages expand in human atrial disease

We collected left atrial tissue from five control participants and seven patients with persistent AFib undergoing heart surgery (table S1) and performed scRNA-seq on the 10X Genomics Chromium platform (Fig. 1A and fig. S1A). A total of 41,609 viable, single cells from both patients with AFib and controls were partitioned into clusters (fig. S1B), which we annotated using known lineage markers (fig. S1C). We identified six major noncardiomyocyte cell types, which included (in order of decreasing frequency) lymphocytes, mononuclear phagocytes and dendritic cells (MP/DCs), endothelial cells, fibroblasts, mural cells, and neutrophils (Fig. 1, B and C; fig. S2; and table S2).

In diseased atria, the MP/DC cluster expanded twofold, whereas endothelia and mural cells decreased in frequency. The other cell classes remained stable (Fig. 1C and table S2). Because the MP/DC cluster increased considerably, we scrutinized the transcriptomes of these cells more closely. We identified genes that were differentially expressed in MP/DCs from AFib and control atria with a pseudo-bulk approach that aggregated cells from the same individual to determine differential gene expression between the two patient groups (*10*). We then performed gene set enrichment analysis (GSEA) (*11*) on the resulting list of differentially expressed genes to define the pathways and functional gene groups that were altered in MP/DCs from patients with AFib. In this disease setting, MP/DCs upregulated gene sets related to the inflammatory response, interferon-γ (IFN-γ) signaling, monocyte recruitment, and cholesterol metabolism (Fig. 1D and data S1). Thus, atrial MP/DCs undergo substantial immunometabolic remodeling during their expansion in patients with AFib.

To gain a broader understanding of the transcriptional programs that are activated or repressed in MP/DCs, we used weighted correlation network analysis to study groups of correlated genes that were differentially expressed between patients and controls (*12*). We identified 24 modules of correlated genes across all participants, three of which were significantly correlated with disease state. Module 1 was up-regulated in AFib and contained inflammatory and profibrotic genes such as *CCR2*, *IL10*, *ITGA9*, *MMP9*, *SPP1*, *TIMP2*, and *VIM* (Fig. 1E and data S2) (*13*–*15*). Overrepresentation analysis confirmed that inflammatory and extracellular matrix–related gene sets characterized module 1 (Fig. 1F and data S2). Module 1 genes *SPP1* and *CCR2* and the cell surface markers *TREM2* and *CD9* were among the top up-regulated genes in MP/DCs from patients with AFib (Fig. 1G). *SPP1* encodes the matricellular protein osteopontin, which participates in bone remodeling and fibrosis by promoting cell survival, adhesion, and migration as well as inflammatory cell activation (*13*). The chemokine receptor CCR2 mediates monocyte egress from the bone marrow (*16*) and is essential for monocyte recruitment (*14*). Gene modules 2 and 3, whose expression was down-regulated in MP/DCs from patients with AFib, included genes involved in protecting against cardiac disease [*SIRT2* (*17*)], cardiac resident macrophage proliferation [*KLF4* (*18*)], and epigenetic constraint of cytokine production [*HDAC2* (*19*) and *ODC1* (*20*)] (Fig. 1E and data S2). These data imply that the atria of patients with AFib have elevated numbers of myeloid cells that assume a more inflammatory phenotype, implicating macrophages in the tissue remodeling that disrupts coordinated excitation of atrial cardiomyocytes. Other leukocytes and stromal

[1]Center for Systems Biology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [2]Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [3]Cardiovascular Research Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [4]Department of Internal Medicine II, University Medical Center Regensburg, Regensburg, Germany. [5]Department of Thoracic and Cardiovascular Surgery, University Hospital Wuerzburg, Wuerzburg, Germany. [6]Department of Pathology, Center for Integrated Diagnostics, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [7]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [8]Radcliffe Department of Medicine, NIHR Biomedical Research Centre, University of Oxford, Oxford, UK. [9]British Heart Foundation Centre of Research Excellence, University of Oxford, Oxford, UK. [10]Institut del Cor Germans Trias i Pujol, CIBERCV, Badalona, Barcelona, Spain. [11]Department of Cardiology and Angiology, University of Giessen/DZHK, Partner Site Rhein-Main, Germany. [12]Division of Cardiac Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [13]Department of Surgery, Yale School of Medicine, New Haven, CT, USA. [14]Cardiovascular Research Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [16]Department of Cardiac Surgery, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [17]Leducq Foundation, Boston, MA, USA. [18]Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard University, Cambridge, MA, USA. [19]Department of Internal Medicine I, University Hospital Wuerzburg, Wuerzburg, Germany.
***Corresponding author. Email: mhulsmans@mgh.harvard.edu (M.H.); naxerova.kamila@mgh.harvard.edu (K.N.); mnahrendorf@mgh.harvard.edu (M.N.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.

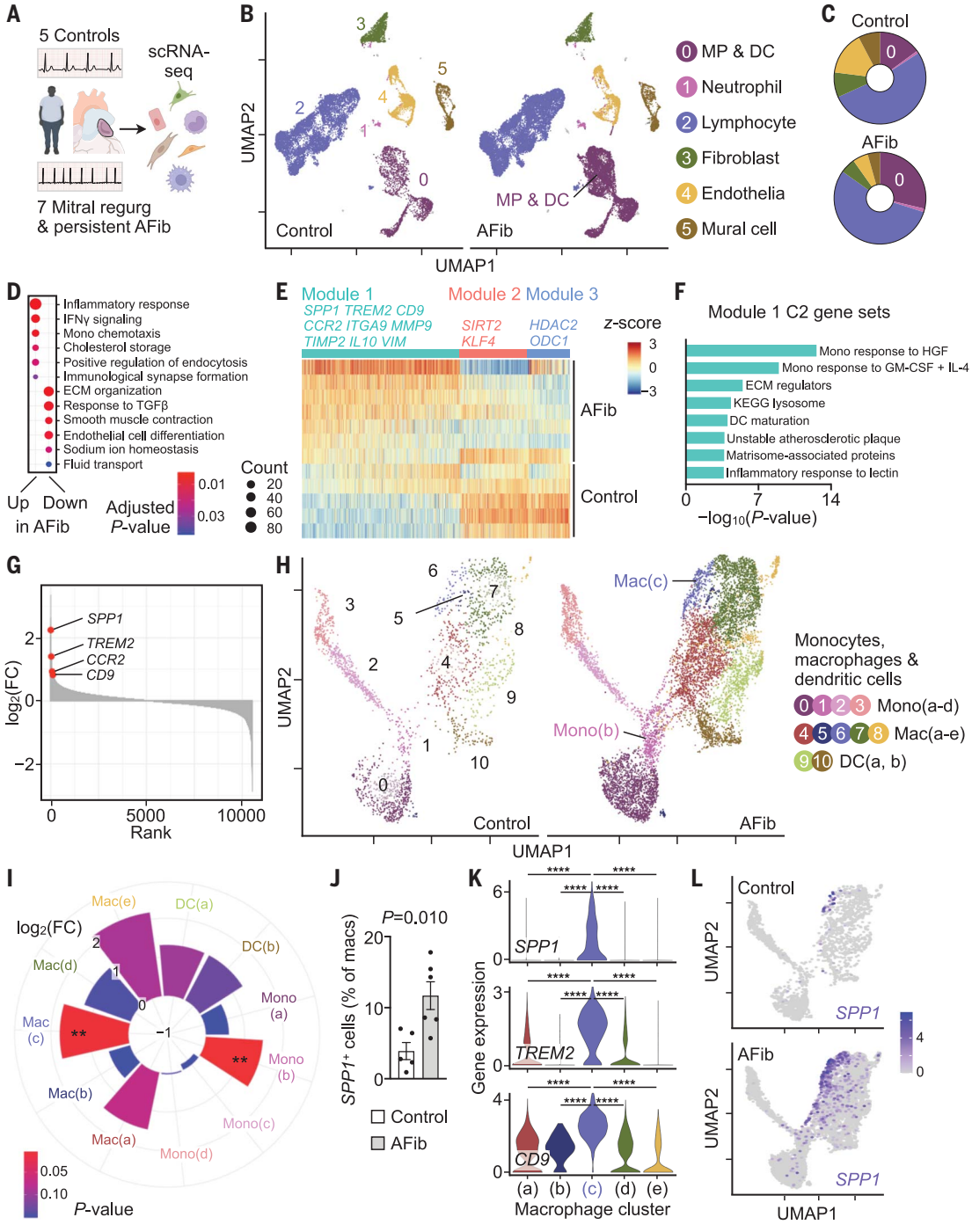


Fig. 1. Single-cell atlas of human atrial disease. (**A**) scRNA-seq was performed in five controls and seven patients with AFib. (**B**) UMAP delineates six major cardiac cell types. (**C**) Distribution of cell populations. Color code corresponds to (B). (**D**) Gene ontology biological process (GOBP) gene sets (*48*, *49*) up- or down-regulated in AFib from GSEA of MP/DC cluster. Circle size denotes number of enriched genes. ECM, extracellular matrix. (**E**) Gene expression in MP/DC cluster represented by *z*-scores of the log-transformed normalized counts. Rows, samples grouped by disease state; columns, genes in the three gene modules significantly associated with disease state. Gene coexpression modules and exemplary genes are labeled. (**F**) Select MSigDB C2 gene sets (*50*) overrepresented in module 1 described in (E). GM-CSF, granulocyte-macrophage colony-stimulating factor; HGF, hepatocyte growth factor; IL-4, interleukin-4; KEGG, Kyoto Encyclopedia of Genes and Genomes. (**G**) Genes in MP/DC cluster ranked according to $\log_2$-fold changes [$\log_2$(FC)] between seven patients with AFib and five controls. (**H**) Subclustering of 10,555 MP/DCs from five controls and seven patients with AFib. (**I**) Difference in abundance of cells belonging to MP/DC subclusters between seven patients with AFib and five control patients. Pie size indicates $\log_2$(FC); color (blue, high; red, low) indicates the *P* value (two-tailed Student's *t* test). $^{**}P < 0.01$. (**J**) Fraction of *SPP1*$^+$ macrophages in controls and patients with AFib. Bar graph shows mean ± SEM, $n = 5$ to 6 per group, two-tailed Student's *t* test. (**K**) Gene expression levels (represented by log-transformed normalized counts) across macrophage clusters; Kruskal-Wallis ($P < 2.2 \times 10^{-16}$) followed by two-tailed Wilcoxon rank sum test. ****Adjusted $P = 2 \times 10^{-16}$. (**L**) *SPP1* expression in MP/DCs from controls and patients with AFib.

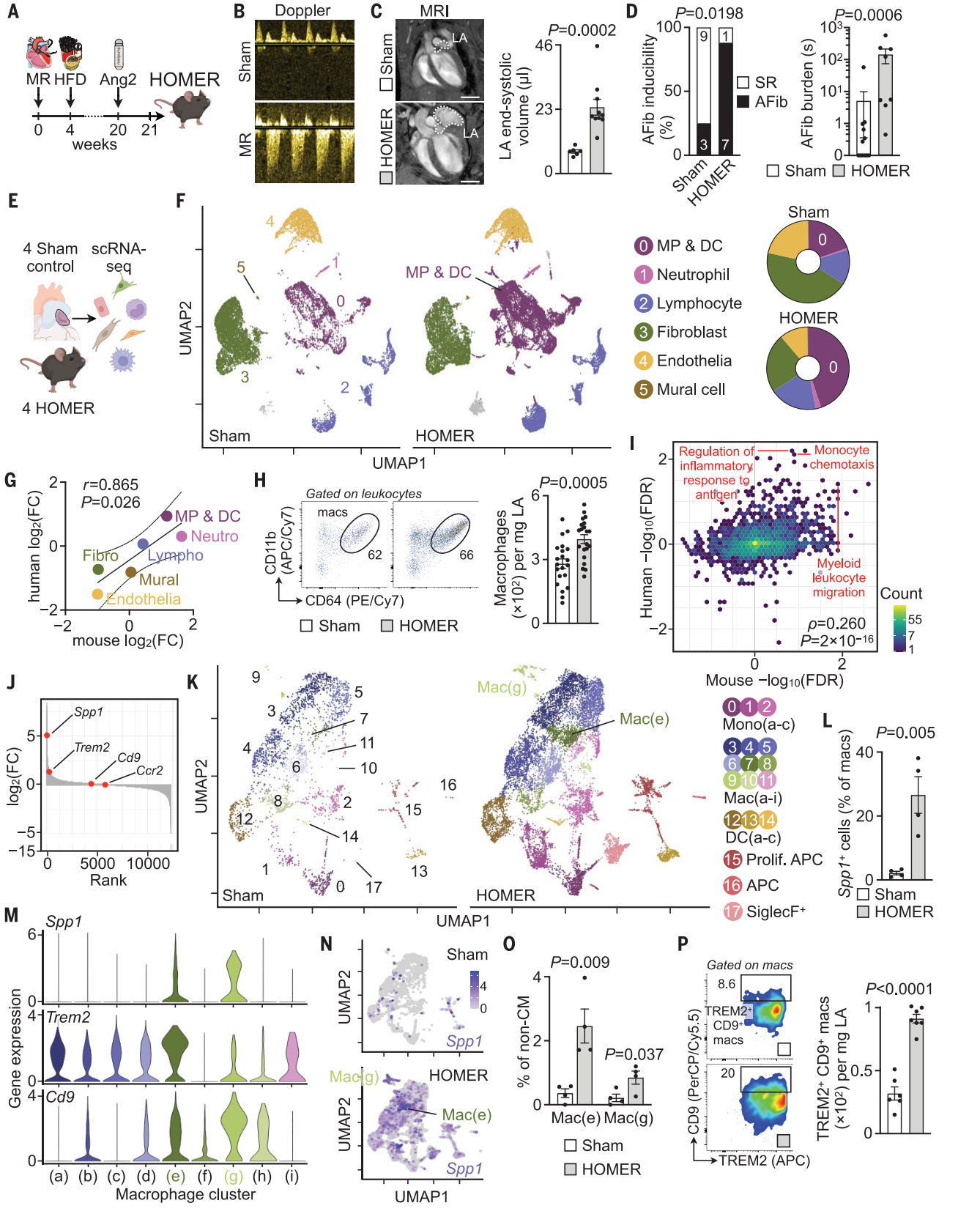


Fig. 2. Single-cell atlas of mouse atria. (**A**) HOMER mice are exposed to mitral valve regurgitation (MR), high-fat diet (HFD), and angiotensin 2 (Ang2). (**B**) Doppler of MR. (**C**) (Left) MRI of left atrium (LA). Scale bars, 3 mm. (Right) LA end-systolic volume, *n* = 6 to 10 per group from two independent experiments, two-tailed Mann-Whitney test. (**D**) AFib inducibility and burden. *n* = 8 to 12 per group from two independent experiments. (Left) Two-sided Fisher's exact test. (Right) Two-tailed Mann-Whitney test. Other parameters are shown in table S7. SR, sinus rhythm. (**E**) scRNA-seq was performed in four

sham and four HOMER mice. (**F**) (Left) UMAP. (Right) Cell population distributions. (**G**) Log_2(FC) of cell abundance in mouse and human atria. (**H**) Flow cytometry of sham and HOMER atria. *n* = 21 per group from five independent experiments, two-tailed Student's *t* test. APC/Cy7, allophycocyanin/cyanine 7; PE/Cy7, phycoerythrin/cyanine 7. (**I**) Scatterplot showing C5 GOBP GSEA species concordance. GSEA is performed on lists of genes ranked according to differential expression in the disease state. $-log_{10}$(FDR) is plotted for mouse (*x* axis) and human (*y* axis). $-log_{10}$(FDR) are negative if the gene set is down-regulated. Yellow indicates high and blue indicates low gene set density. (**J**) Genes in MP/DC cluster ranked according to log_2(FC) between four HOMER and four sham mice. The *y* axis was broken at −5 for displaying the top down-regulated genes. (**K**) Subclustering of MP/DCs from four sham and four HOMER mice. APC, antigen-presenting cell. (**L**) Atrial *Spp1*$^+$ macrophages by scRNA-seq. *n* = 4 per group, two-tailed Student's *t* test. (**M**) Gene expression levels (represented by log-transformed normalized counts) across macrophage clusters. (**N**) *Spp1* expression in MP/DCs from sham and HOMER mice. (**O**) *Spp1*$^+$ subpopulations in sham and HOMER mice. *n* = 4 per group, two-tailed Student's *t* test. non-CM, noncardiomyocyte. (**P**) Flow cytometry of atrial macrophages in sham and HOMER mice. *n* = 6 to 7 per group from two independent experiments, two-tailed Student's *t* test. PerCP/Cy5.5, peridinin-chlorophyll-protein/cyanine 5.5. All bar graph data are mean ± SEM with individual values for data distribution.

cells also altered their gene expression (fig. S3 and data S3 and S4). Fibroblasts, for example, responded by up-regulating inflammatory pathways and matrix production (fig. S3C). In endothelial and mural cells, we observed up-regulation of genes indicative of ischemia and vascular remodeling (fig. S4).

Depending on their origin and microenvironment, macrophages and their subsets take on phenotypes that enable a range of diverging and even opposing functions (*3*, *21*). To identify cell types that contribute to atrial remodeling, we examined atrial myeloid cells in detail. We reclustered MP/DCs, choosing a resolution that maximized the number of differentially expressed genes between subclusters (fig. S5, A and B). Reclustering separated three monocyte subsets, an established number in human blood (*21*), and a fourth inflammatory monocyte cluster whose expression profile (fig. S5C and data S5) and uniform manifold approximation and projection (UMAP) position adjacent to macrophages (Fig. 1H) suggested an intermediate phenotype during monocyte-to-macrophage differentiation. At this resolution, we could distinguish five macrophage clusters (Fig. 1H, figs. S5C and S6, and data S6) and two DC clusters (Fig. 1H and fig. S5C). The inflammatory mono(b) cluster and the *SPP1*$^+$ mac(c) cluster expanded 2.8- and 3.4-fold, respectively, in diseased atria (Fig. 1I). The frequency of *SPP1*$^+$ atrial cells increased in patients with AFib when compared with non-AFib controls (Fig. 1J). The expanding mac(c) cluster was the dominant *SPP1* source in atria of patients with AFib (Fig. 1, K and L); other immune and stromal cells did not express *SPP1* at high levels (fig. S7). *TREM2* and *CD9*, which encode proteins that can be detected by cell surface staining (*22*), further distinguished the *SPP1*$^+$ mac(c) cluster from other cells (Fig. 1K and fig. S7). RNA in situ hybridization confirmed that macrophages express *SPP1* in atria of patients with AFib (fig. S8, A and B). Once secreted by macrophages, SPP1 associates with extracellular matrix (*13*).

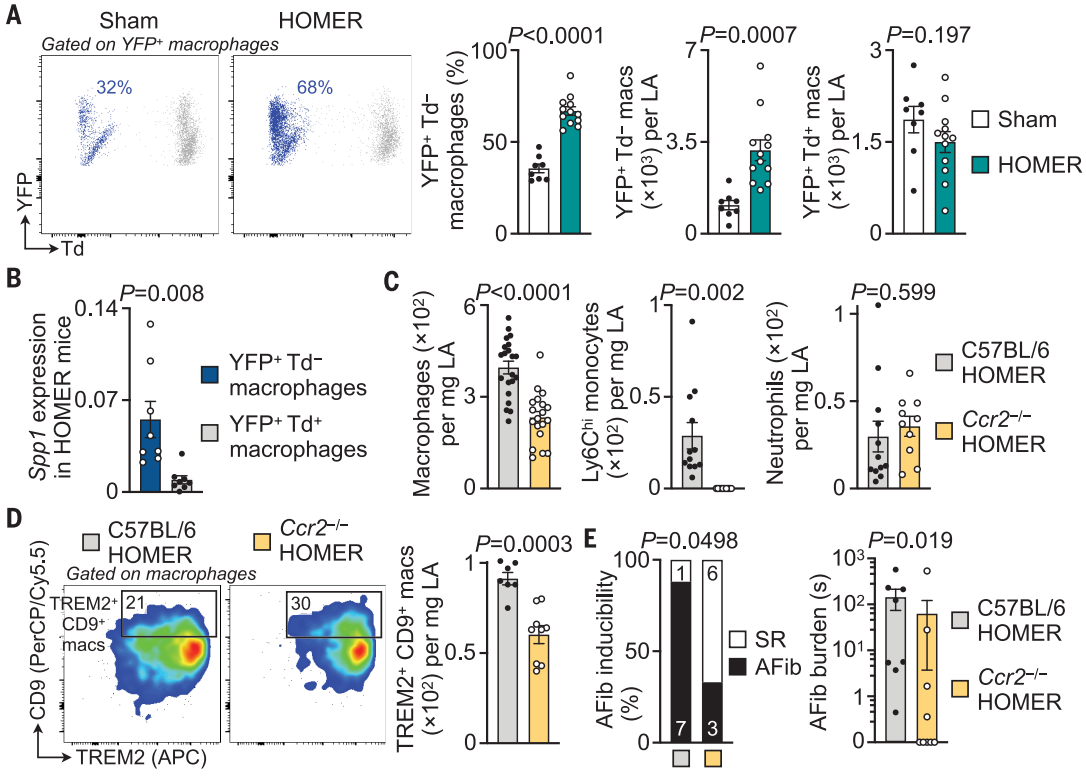


Fig. 3. CCR2-dependent monocyte recruitment contributes to left atrial macrophage expansion and atrial disease. (**A**) Flow cytometric quantification of recruited (YFP$^+$ Td$^-$) and locally sourced (YFP$^+$ Td$^+$) left atrial macrophages in *Cx3cr1*$^{CreER/+}$*Ai9*$^{fl/+}$ sham and HOMER mice. *n* = 8 to 12 per group from two independent experiments, two-tailed Student's *t* test. Td, tdTomato; YFP, yellow fluorescent protein. (**B**) Relative *Spp1* expression levels by qPCR in recruited (YFP$^+$ Td$^-$) and locally sourced (YFP$^+$ Td$^+$) left atrial macrophages in *Cx3cr1*$^{CreER/+}$*Ai9*$^{fl/+}$ HOMER mice. *n* = 8 per group from two independent experiments, two-tailed Wilcoxon matched-pairs signed-rank test. (**C**) Flow cytometric quantification of myeloid cell populations in left atrial tissues from C57BL/6 and *Ccr2*$^{-/-}$ HOMER mice. Macrophages: 21 C57BL/6 and 19 *Ccr2*$^{-/-}$ HOMER mice from five independent experiments; monocytes and neutrophils: 12 C57BL/6 and 10 *Ccr2*$^{-/-}$ HOMER mice from three independent experiments; two-tailed Student's *t* test. (**D**) Flow cytometric quantification of TREM2$^+$ CD9$^+$ left atrial macrophages in C57BL/6 and *Ccr2*$^{-/-}$ HOMER mice. *n* = 7 to 9 per group from two independent experiments, two-tailed Student's *t* test. (**E**) AFib inducibility and burden in C57BL/6 and *Ccr2*$^{-/-}$ HOMER mice by means of electrophysiological (EP) study. *n* = 8 to 9 per group from two independent experiments. (Left) Two-sided Fisher's exact test. (Right) Two-tailed Mann-Whitney test. Other EP parameters are shown in table S8. All bar graph data are mean ± SEM with individual values for data distribution.

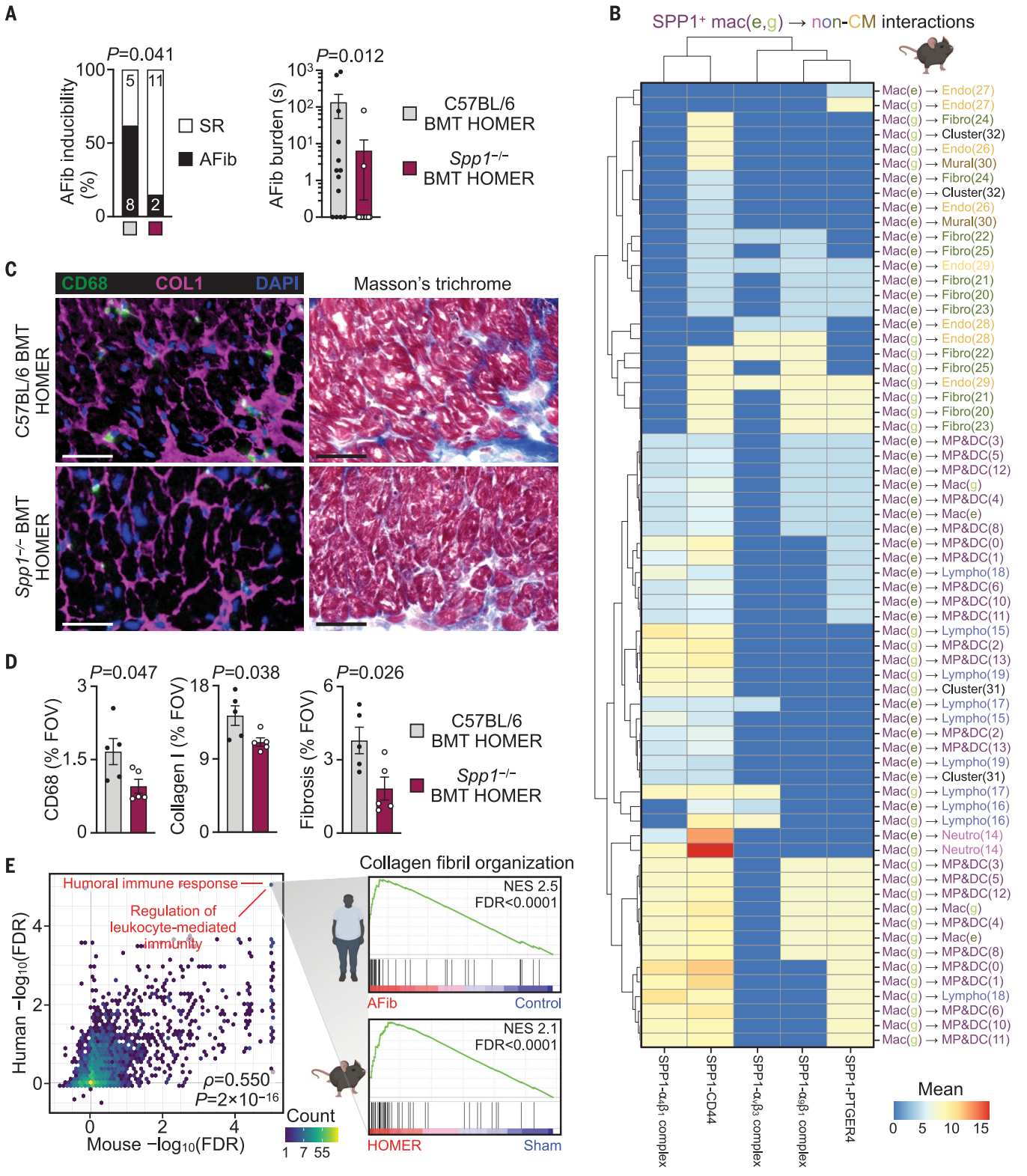


Fig. 4. Macrophage-derived SPP1 promotes atrial disease. (**A**) EP study of AFib inducibility and burden in C57BL/6 BMT HOMER and *Spp1*$^{-/-}$ BMT HOMER mice. *n* = 13 per group from three independent experiments. (Left) Two-sided Fisher's exact test. (Right) Two-tailed Mann-Whitney test. Other EP parameters are shown in table S9. (**B**) Heatmap of SPP1-receptor interactions between SPP1^{+} macrophages [mac(e) and mac(g)] and other non-CMs in four HOMER mice. SPP1 and cognate receptors are shown on the *x* axis. Cell populations that express *Spp1* and the receptor are shown on the *y* axis and are color-coded according to the major cardiac cell types identified in Fig. 2F. Color of scale bar (red, high; blue, low) indicates average *Spp1* and receptor expression levels in their respective interacting subpopulations (represented by mean normalized counts) if the enrichment of an interacting SPP1-receptor pair in the given interacting subpopulations was statistically significant ($P < 0.05$). Nonsignificant interactions were assigned a value of 0. (**C**) Representative immunofluorescent

staining of macrophages (CD68, green), collagen deposition (COL1, magenta), and nuclei [4′,6-diamidino-2-phenylindole (DAPI), blue] and histological staining of fibrosis (Masson's trichrome staining) in left atrial tissue from C57BL/6 BMT HOMER and *Spp1*$^{-/-}$ BMT HOMER mice. Scale bars, 50 μm. (**D**) Quantification of CD68^{+}, COL1^{+}, and fibrotic area in left atrial tissue from C57BL/6 BMT HOMER and *Spp1*$^{-/-}$ BMT HOMER mice. Bar graphs show percentage of positive staining per field of view (FOV); n = 5 per group from two independent experiments, two-tailed Student's t test. (**E**) Scatterplot showing concordance between C5 GOBP GSEA results in human and mouse fibroblasts in AFib and HOMER (as in Fig. 2I). (Inset) Detailed GSEA results for gene set "GOBP collagen fibril organization" in human and mouse fibroblast clusters. NES, normalized enrichment score. All bar graph data are mean ± SEM with individual values for data distribution.

Accordingly, immunofluorescence microscopy identified expanding colocalization of SPP1 protein with collagen in patients with AFib (fig. S8, C and D).

To explore atrial macrophage expansion in a larger cohort, we used immunofluorescence histology on left atrial appendages of 108 patients with AFib and/or mitral regurgitation and 41 controls in sinus rhythm (table S3). Atrial macrophages and their monocyte-derived CCR2^{+} subset expanded most in patients with AFib and mitral regurgitation (fig. S8, E and F). In patients who had mitral regurgitation without AFib or AFib without mitral regurgitation, macrophages expanded to a lesser degree (fig. S8, E and F). In patients who had AFib without mitral regurgitation, AFib chronicity was less pronounced (table S3), which may explain why we observed fewer atrial macrophages in this cohort.

A mouse model of atrial disease

To dissect immune pathways involved in human AFib mechanistically, we used a mouse model that combined hypertension, obesity, and mitral valve regurgitation (HOMER) (Fig. 2A). Snaring the mitral valve to the ventricular wall produced mitral valve regurgitation (Fig. 2B). Mice consumed an obesogenic diet for 4 months, mirroring the risk that obesity engenders for AFib (*23*). Hypertension, also a risk factor for AFib, was induced with angiotensin 2 (fig. S9). The HOMER model elicited left atrial enlargement (Fig. 2C) commonly found in patients with AFib (*24*). HOMER mice developed increased AFib inducibility and burden compared with those of sham-operated normotensive control mice that consumed standard chow (hereafter referred to as controls) (Fig. 2D and fig. S10). Among individual HOMER components, obesity elicited the highest AFib inducibility, whereas mitral valve regurgitation enlarged atria the most (fig. S11). Spontaneous AFib episodes, which were monitored with electrocardiogram telemeters, were absent in HOMER mice. In HOMER hearts, left ventricles hypertrophied and dilated mildly, and left ventricular systolic function was slightly reduced (fig. S12A and table S4). Blood leukocyte counts, including Ly6C^{hi} monocytes, rose in HOMER mice (fig. S12, B and C). Histological analysis revealed more numerous macrophages, fibrosis, and collagen deposition in HOMER atria compared with those in controls (fig. S12, D and E), results that align with the fibrosis observed in patients with AFib (*25*) and the effects of hypertension and metabolic stress on the ventricles (*15*, *26*, *27*). Hypertension and mitral valve regurgitation most clearly increased atrial macrophage numbers and fibrosis (fig. S13).

Replicating our human scRNA-seq pipeline by using left atrial tissues dissected from four HOMER mice and four control animals (Fig. 2E and fig. S14A), we successfully obtained 40,112 viable single cells, among which we identified the same cell types as in human atria (Fig. 2F; fig. S14, B and C; and table S5). The cellular composition changes in diseased atria correlated strongly between both species [Pearson's correlation coefficient (r) = 0.87] (Fig. 2G). We observed a twofold expansion of MP/DCs in HOMER mice similar to that in the human disease setting (Fig. 2F and table S5). Flow cytometry confirmed the HOMER-induced changes in the abundance of macrophages (Fig. 2H and fig. S15, A to C). Cardiac macrophage subsets can be distinguished by expression of CCR2 and major histocompatibility complex II (MHCII) (*28*, *29*). In atria of HOMER mice, the number of CCR2^{+} macrophages and monocytes rose (fig. S15, D and E).

We next examined the gene expression changes that occurred in MP/DCs of HOMER mice in relation to our human data. As before, we first performed GSEA on differentially expressed genes from a pseudo-bulk analysis comparing MP/DCs isolated from atria of HOMER or control mice. We then plotted the significance and direction of change for all Molecular Signatures Database (MSigDB) C5 (Gene Ontology) gene sets against the corresponding values in humans and found significant concordance between the two GSEA data sets [Spearman's rank correlation coefficient (ρ) = 0.26, $P < 2 \times 10^{-16}$] (Fig. 2I and data S7), indicating that the HOMER model recapitulated the pathway-level gene expression changes observed in patients. As in humans, *Spp1* was among the top up-regulated genes in MP/DCs of HOMER atria (Fig. 2J). We then reclustered MP/DCs to obtain a higher-resolution view of these cells, choosing a resolution that separated three monocyte clusters [a number previously described in mouse blood (*30*)], nine macrophage clusters, three DC clusters, two antigen-presenting cell clusters, and one SiglecF^{+} cell cluster (Fig. 2K). *Spp1*$^{+}$ cell frequencies increased in HOMER atria (Fig. 2L), a rise driven by the macrophage subpopulations mac(e) and mac(g) and not by other immune or stromal cells (Fig. 2, M and N, and fig. S16, A and B). In close concordance with our human findings, *Spp1*$^{+}$ mac(e) and mac(g) clusters coexpressed *Trem2* and *Cd9* (Fig. 2M) and expanded in HOMER mice (Fig. 2, O and P, and fig. S16, C and D). Comparing MP/DC clusters across the two species (using the top 100 genes that distinguished each subcluster), we found that each human MP/DC cluster significantly matched at least one mouse cluster and that the human *SPP1*$^{+}$ cluster mac(c) was similar to the mouse *Spp1*$^{+}$ cluster mac(e) (fig. S17). This strong interspecies correlation is likely caused by similar risk factors in human patients and HOMER mice.

We next assayed contributions of individual HOMER components. *Ccr2* was expressed highest in HOMER atria but also increased in hypertensive mice (fig. S18A). Atrial expression of *Csf1*, chemokines, adhesion molecules, inflammatory markers, and matrix increased the most in HOMER mice (fig. S18, B to F). *Spp1* expression increased in HOMER mice but not after exposure to single components (fig. S18G). Thus, HOMER components have additive (for example, *Ccr2* expression) and some synergistic effects (for example, *Spp1* expression). Two of three HOMER components elicited intermediate atrial remodeling and *Spp1* expression (fig. S19). Although combining all HOMER components reflected the risk factors present in our patient cohort most faithfully, further studies of single conditions (for example, metabolic consequences of high-fat diet on conducting cells) are warranted.

We then contrasted left atrial gene expression changes to the left ventricular myocardium. Unlike the atria, we observed no significant expansion of ventricular leukocyte frequencies in HOMER mice (fig. S20, A to C, and table S6). The decline of fibroblast frequencies that occurred in HOMER atria was absent in the ventricles (fig. S20, A to C, and table S6). Compared with ventricular cells, more genes changed in atrial myeloid cells and fibroblasts. For example, there were 1982 atrial but only 19 ventricular differentially expressed genes [with false discovery rate (FDR) < 0.05] in MP/DCs (with nine differentially regulated genes shared between the atria and ventricles) (fig. S20, D and E, and data S8). This pattern was inversed for endothelial cells, which had more differentially expressed genes in the ventricles than in the atria (fig. S20, D and E, and data S8). A principal components analysis revealed that in controls, atrial MP/DCs were located at a large distance from ventricular MP/DCs. The distance between atrial and ventricular MP/DCs

was less pronounced in HOMER mice (fig. S20F). We hypothesize that HOMER-induced activation of inflammatory programs equalizes some of the differences normally imprinted by the atrial versus the ventricular environment. That the HOMER model affects left ventricular leukocytes to a lesser degree than in the left atria corresponds with the more severe atrial than ventricular dilation observed with magnetic resonance imaging (MRI) (Fig. 2C and fig. S12A). The ventricular data reveal that hypertension, high-fat diet, and mitral valve disease affect not only atria but the entire heart, a notion reflected by the clinical association of AFib with heart failure (*31*).

The generation of the HOMER mouse and its scRNA-seq validation against patients with AFib provided us with a tool to assess whether immune cell expansion merely associates with—or alternatively, contributes to—AFib. We used the HOMER mouse for two follow-up studies that tested causal implications for key observations in patients with AFib: (i) expansion of inflammatory atrial macrophages and (ii) increased atrial SPP1 expression by those macrophages.

CCR2-dependent macrophage recruitment promotes AFib

Macrophage ontogeny often dictates function (*32*). For example, steady-state resident macrophages in the heart and brain are distinctly noninflammatory (*33*), whereas recruited infarct or stroke macrophages are inflammatory and damage tissues (*34*). We consequently sought to investigate macrophage origins in normal and diseased atria. Because atrial macrophages expand in diseased atria of HOMER mice (Fig. 2, F and H, and table S5) and enrich for the chemokine receptor CCR2 (fig. S15D), which recruited macrophages express (*34*), we combined the HOMER procedure with genetic fate mapping in *Cx3cr1*$^{CreER/+}$*Ai9*$^{fl/+}$ mice (fig. S21A). Although atrial macrophages were more likely to be derived from circulating monocytes than ventricular macrophages, most atrial macrophages were not sourced from monocytes in steady state (fig. S21B). In HOMER mice, recruitment expanded so that 67% of atrial macrophages derived from monocytes, a significant increase from the steady state (36%) (Fig. 3A). In line with increased atrial expression of *Csf1* (fig. S18B) and up-regulation of the "Leukocyte proliferation" gene set by MP/DCs (fig. S21C), macrophage proliferation increased in HOMER atria (fig S21, D and E). Recruited macrophages in atria of HOMER mice were the primary source of *Spp1*, whereas locally sourced macrophages did not express this gene at high levels (Fig. 3B).

We used the HOMER procedure in *Ccr2*$^{-/-}$ mice to test the relevance of expanded macrophage recruitment for AFib. Blocking monocyte migration in *Ccr2*$^{-/-}$ HOMER mice decreased blood monocyte, atrial Ly6C^{hi} monocyte, and macrophage expansion when compared with C57BL/6 HOMER mice, whereas other cell densities and blood pressure were unaffected (Fig. 3C and fig. S22, A to D). TREM2^{+} CD9^{+} macrophages were also lower in *Ccr2*$^{-/-}$ HOMER mice (Fig. 3D). Cardiac MRI revealed smaller left atrial volumes in *Ccr2*$^{-/-}$ HOMER mice (fig. S22E). *Ccr2*$^{-/-}$ HOMER mice exhibited lower AFib inducibility and AFib burden than those of C57BL/6 HOMER mice (Fig. 3E). Thus, macrophage recruitment expands during atrial remodeling, recruited macrophages highly express *Spp1*, and inhibiting macrophage recruitment reduces AFib.

To test whether CCR2 inhibition could alleviate established disease, we treated mice with the CCR2 antagonist CCL2-Fc (*35*) for 4 weeks after the last HOMER treatment component (fig. S23A). CCR2 inhibition reduced fibrosis in atria of HOMER mice (fig. S23, B and C). AFib inducibility and burden were lower in the treatment group, although this effect was not statistically significant (fig. S23D).

Macrophage-derived SPP1 is a pleiotropic AFib catalyst

Given that *Spp1* is among the most up-regulated macrophage genes in AFib, we examined the causal role of this matricellular protein. We transplanted bone marrow from *Spp1*$^{-/-}$ mice into wild-type recipients (*Spp1*$^{-/-}$ BMT HOMER) to delete *Spp1* in recruited macrophages (fig. S24, A and B). Wild-type recipients that received wild-type bone marrow and then underwent HOMER (C57BL/6 BMT HOMER) served as controls. Blood pressure and leukocyte counts were similar in both cohorts (fig. S24, C to E). Left atrial and ventricular volumes were preserved and left ventricular hypertrophy was lessened in mice in which *Spp1* was deleted (fig. S24F). Deleting *Spp1* in bone marrow–derived cells reduced AFib inducibility and AFib burden (Fig. 4A).

To outline mechanisms by which osteopontin exacerbates AFib, we investigated the crosstalk between *Spp1*-expressing macrophages and other atrial cells, using CellPhoneDB (*36*). In HOMER mice and patients with AFib, paracrine and autocrine interactions were found between macrophage-derived SPP1 and cognate SPP1 receptors expressed by immune and stromal cells (Fig. 4B and fig. S25). In HOMER mice, CellPhoneDB analysis implied SPP1^{+} macrophage interactions with client cells via three integrins ($\alpha_4\beta_1$, $\alpha_v\beta_3$, and $\alpha_9\beta_1$), CD44, and prostaglandin E_2 receptor 4 (EP_4/PTGER4). These interactions, which included signaling to various macrophage and fibroblast clusters, suggested that SPP1 acts as a pleiotropic amplifier for atrial inflammation and fibrosis, which in turn contribute to an AFib substrate (*37*).

In support of this idea, the histology of atria collected from *Spp1*$^{-/-}$ BMT HOMER mice showed fewer CD68^{+} macrophages when compared with that of C57BL/6 BMT HOMER mice (Fig. 4, C and D), indicating that SPP1 may influence atrial macrophage supply. Atrial monocyte numbers were unchanged in *Spp1*$^{-/-}$ BMT HOMER mice (fig. S26A), making a role of SPP1 in monocyte recruitment unlikely. TUNEL (terminal deoxynucleotidyl transferase–mediated deoxyuridine triphosphate nick end labeling) and CD68 staining of atria indicated that macrophage death was also unchanged in the cohort in which *Spp1* was deleted (fig. S26B). However, macrophage proliferation, which is up-regulated by HOMER conditions (fig. S21, D and E), declined in *Spp1*$^{-/-}$ BMT HOMER compared with that in C57BL/6 BMT HOMER mice (fig. S26, C and D). Thus, SPP1 augments atrial inflammation through increased macrophage proliferation.

Matrix production by fibroblasts increased in HOMER mice (S26, E and F), and immunoreactive α smooth muscle actin (αSMA) staining of HOMER atria documented an expansion of activated fibroblasts, which were reduced in *Spp1*$^{-/-}$ BMT HOMER mice (fig. S26, G to J). Atrial fibroblasts isolated from *Spp1*$^{-/-}$ BMT HOMER mice expressed less collagen than those from C57BL/6 BMT HOMER (fig. S26K). Furthermore, we detected reductions in collagen I deposition and Masson's trichrome staining in atria of *Spp1*$^{-/-}$ BMT HOMER mice (Fig. 4, C and D), which points to reduced fibroblast matrix production if these cells do not sense SPP1. This interaction may be particularly relevant because integrins—which, according to our CellPhoneDB analysis, fibroblasts use to sense SPP1—regulate fibroblast migration, proliferation, and activation, processes that precede increased tissue fibrosis (*38*, *39*).

Comparing left atrial cells from *Spp1*$^{-/-}$ BMT HOMER with those from C57BL/6 BMT HOMER mice by means of scRNA-seq (fig. S27, A and B), we identified fibroblast clusters that expressed collagens at high levels (fig. S27, C to E). For these clusters, we pursued GSEA to examine through which pathways SPP1 activates atrial fibroblasts (fig. S27F). Fibroblast clusters from C57BL/6 BMT HOMER mice up-regulated *Tgfb1* gene sets. Inflammatory activation of fibroblasts by SPP1 was supported by the up-regulation of inflammatory gene sets in fibroblasts from C57BL/6 BMT HOMER when compared with *Spp1*$^{-/-}$ BMT HOMER mice (fig. S27F). Thus, SPP1 participates in inflammatory fibroblast activation partially through transforming growth factor–β (TGF-β) pathways. In atrial tissue from patients with AFib, SPP1 colocalization with collagen 1A1 expanded (fig. S8, C and D), lending further credence to SPP1 promotion of AFib through enhanced fibrosis.

To ascertain whether the transcriptional changes that occurred in atrial fibroblasts of HOMER mice are representative of alterations

in humans, we compared human and mouse fibroblast transcriptomes. We first performed GSEA (again using all C5 ontology MSigDB gene sets) on differentially expressed genes from a pseudo-bulk analysis comparing all fibroblast clusters isolated from atria of patients with AFib versus humans without AFib, as well as HOMER mice versus control mice. We then plotted the significance and direction of change for all gene sets for both species. Both GSEA data sets correlated strongly (Spearman's $\rho = 0.55$, $P < 2 \times 10^{-16}$) with gene sets related to immune response and collagen organization enriched in diseased atria (Fig. 4E and data S9). This pathway-level interspecies correlation further bolstered the physiological relevance of the HOMER model while also underscoring inflammation and fibrosis as central processes in AFib.

Discussion

In this study, we comprehensively assayed atria from patients and mice and documented large-scale shifts in the cellular frequencies and phenotypes during AFib. Macrophages expand more than any other cell type in AFib. Experiments in mice, in which either *Ccr2* or *Spp1* was deleted, demonstrated that targeting macrophages strengthens resilience against AFib (fig. S28).

In human and mouse atria, *SPP1* was among the top up-regulated macrophage genes. This matricellular protein, which increases in the blood of patients with AFib (*40*), stabilizes collagen (*41*, *42*) and promotes fibrosis in hypertension (*43*, *44*). We speculate that macrophage-derived SPP1 (*45*) augments atrial tissue heterogeneity and hinders conduction between cardiomyocytes. SPP1 signals to most other atrial cell classes, enhancing inflammation and fibrosis through interaction with various cell surface receptors. In the ventricular myocardium and in other organs, such signaling promotes myeloid cell activation and migration as well as extracellular matrix production by fibroblasts (*13*, *15*, *43*, *44*). Thus, targeting SPP1 may have several distinctive effects that together preserve atrial conduction. The expanding numbers and altered phenotypes of atrial macrophages may have other disease-promoting consequences. For example, a loss of resident macrophages' relationships with cardiomyocytes (*46*, *47*) may compromise atrial conduction.

Going forward, therapeutic strategies could target recruitment of inflammatory macrophages (CCR2) or macrophage-derived signals (SPP1). Macrophage-targeted therapy for atrial cardiomyopathy is most likely to succeed with concomitant normalization of risk factors—for example, through surgical valve repair, weight loss, and blood pressure management. Our data centrally position macrophages as targets for immunotherapy in patients with AFib.

REFERENCES AND NOTES

1. M. Litviňuková *et al*., *Nature* **588**, 466–472 (2020).
2. J. A. Nicolás-Ávila, A. Hidalgo, I. Ballesteros, *J. Leukoc. Biol.* **104**, 743–756 (2018).
3. E. Forte, M. B. Furtado, N. Rosenthal, *Nat. Rev. Cardiol.* **15**, 601–616 (2018).
4. E. J. Benjamin *et al*., *Circulation* **139**, e56–e528 (2019).
5. T. Yamashita *et al*., *Circ. J.* **74**, 262–270 (2010).
6. N. Smorodinova *et al*., *PLOS ONE* **12**, e0172691 (2017).
7. C. Yao *et al*., *Circulation* **138**, 2227–2242 (2018).
8. Y. K. Iwasaki, K. Nishida, T. Kato, S. Nattel, *Circulation* **124**, 2264–2274 (2011).
9. V. Rudolph *et al*., *Nat. Med.* **16**, 470–474 (2010).
10. M. D. Robinson, D. J. McCarthy, G. K. Smyth, *Bioinformatics* **26**, 139–140 (2010).
11. A. Subramanian *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
12. P. Langfelder, S. Horvath, *BMC Bioinformatics* **9**, 559 (2008).
13. N. G. Frangogiannis, *Physiol. Rev.* **92**, 635–688 (2012).
14. I. F. Charo, R. M. Ransohoff, *N. Engl. J. Med.* **354**, 610–621 (2006).
15. M. Hulsmans *et al*., *J. Exp. Med.* **215**, 423–440 (2018).
16. N. V. Serbina, E. G. Pamer, *Nat. Immunol.* **7**, 311–317 (2006).
17. X. Tang *et al*., *Circulation* **136**, 2051–2067 (2017).
18. X. Liao *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4661–E4669 (2018).
19. Q. Zhang *et al*., *Nature* **525**, 389–393 (2015).
20. D. M. Hardbower *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, E751–E760 (2017).
21. E. K. Koltsova, C. C. Hedrick, K. Ley, *Curr. Opin. Lipidol.* **24**, 371–380 (2013).
22. P. Ramachandran *et al*., *Nature* **575**, 512–518 (2019).
23. R. R. Huxley *et al*., *Circulation* **123**, 1501–1508 (2011).
24. T. S. Tsang *et al*., *Mayo Clin. Proc.* **76**, 467–475 (2001).
25. A. Xintarakou, S. Tzeis, S. Psarras, D. Asvestas, P. Vardas, *Europace* **22**, 342–351 (2020).
26. J. D. Schilling, H. M. Machkovech, A. H. Kim, R. Schwendener, J. E. Schaffer, *Am. J. Physiol. Heart Circ. Physiol.* **303**, H1366–H1373 (2012).
27. L. Wang *et al*., *Eur. Heart J.* **39**, 1818–1831 (2018).
28. S. Epelman *et al*., *Immunity* **40**, 91–104 (2014).
29. G. Bajpai *et al*., *Nat. Med.* **24**, 1234–1245 (2018).
30. L. Ziegler-Heitbrock *et al*., *Blood* **116**, e74–e80 (2010).
31. R. Gopinathannair *et al*., *Circ. Arrhythm. Electrophysiol.* 10.1161/HAE0000000000000078 (2021).
32. C. Blériot, S. Chakarov, F. Ginhoux, *Immunity* **52**, 957–970 (2020).
33. F. F. Hoyer *et al*., *Immunity* **51**, 899–914.e7 (2019).
34. S. Epelman, K. J. Lavine, G. J. Randolph, *Immunity* **41**, 21–35 (2014).
35. C. D. Paavola *et al*., *J. Biol. Chem.* **273**, 33157–33165 (1998).
36. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, *Nat. Protoc.* **15**, 1484–1506 (2020).
37. M. Harada, S. Nattel, *Card. Electrophysiol. Clin.* **13**, 25–35 (2021).
38. K. Graf *et al*., *Hypertension* **35**, 978–984 (2000).
39. V. Sarrazy *et al*., *Cardiovasc. Res.* **102**, 407–417 (2014).
40. J. Molvin *et al*., *Open Heart* **7**, e001190 (2020).
41. L. Liaw *et al*., *J. Clin. Invest.* **101**, 1468–1478 (1998).
42. S. R. Künzel *et al*., *Circ. Res.* **129**, 804–820 (2021).
43. A. R. Collins *et al*., *J. Am. Coll. Cardiol.* **43**, 1698–1705 (2004).
44. Y. Matsui *et al*., *Hypertension* **43**, 1195–1201 (2004).
45. B. López *et al*., *Cardiovasc. Res.* **99**, 111–120 (2013).
46. J. A. Nicolás-Ávila *et al*., *Cell* **183**, 94–109.e23 (2020).
47. M. Hulsmans *et al*., *Cell* **169**, 510–522.e20 (2017).
48. M. Ashburner *et al*., *Nat. Genet.* **25**, 25–29 (2000).
49. S. Carbon *et al*., *Nucleic Acids Res.* **49** (D1), D325–D334 (2021).
50. A. Liberzon *et al*., *Cell Syst.* **1**, 417–425 (2015).

ACKNOWLEDGMENTS

We thank A. Bard and N. Higgins for pathology assistance and K. Joyes for editing the manuscript. The authors also thank the HSCI-CRM Flow Cytometry Core for assistance with cell sorting and the BWH Specialized Histopathology Services Core for histology support. We acknowledge Servier Medical Art (https://smart.servier.com) and BioRender (IH23GLWC6I) for cartoon components. **Funding:** This work was supported by National Institutes of Health grants DK040561 (M.H.), HL155097 (M.H.), HL149647 (M.H.), HL142494 (K.N. and M.N.), HL139598 (M.N.), CA225655 (K.N.), HL007604 (A.B.), HL158040 (F.E.P.), HL092577 (P.T.E.), HL157635 (P.T.E.), and HL139731 (P.T.E.); American Heart Association grants 19CDA34490005 (M.H.) and 18SFRN34110082 (P.T.E.); Deutsche Forschungsgemeinschaft (DFG) #453989101 SFB 1525 Mercator fellow (M.N.); DFG (M.J.S. and S.S.); British Heart Foundation (A.L. and B.C.); NIHR Oxford Biomedical Research Centre (A.L. and B.C.); and European Union MAESTRIA 965286 (P.T.E.). **Author contributions:** M.H., M.J.S., and I.L. conceived the study; designed, performed, and analyzed experiments; interpreted data; and made the figures. A.B., Y.I., C.V., A.P., M.Y., J.G., N.M., H.S., N.K., F.E.P., D.R., and G.R.W. designed, performed, and analyzed experiments. M.J.S. established and validated the MR surgery. S.P., D.K., C.B., U.G., J.K.L., R.N.M., A.L., B.C., O.I.-E., A.B.-G., S.S., C.S.O., R.N.P., G.Te., G.To., and S.M. provided human samples. J.C.A., R.W., F.K.S., D.J.M., and P.T.E. discussed results and strategy. K.N. and M.N. conceived and directed the study. M.H. and M.N. wrote the manuscript with input from all authors. **Competing interests:** M.H., M.J.S, I.L., N.M., K.N., and M.N. are inventors on a patent application related to this work (US patent application no. 63/501,286). J.C.A. serves on the scientific advisory board of Cellestia and is a consultant to Remix Therapeutics and Ayala Pharmaceuticals. P.T.E. receives funds or material research support from Bayer AG, IBM Research, Bristol Myers Squibb, Pfizer, MyoKardia, and Novartis. M.N. has received funds or material research support from Alnylam, Biotronik, CSL Behring, GlycoMimetics, GSK, Medtronic, Novartis, and Pfizer, as well as consulting fees from Biogen, Gimv, IFM Therapeutics, Molecular Imaging, Sigilon, Verseau Therapeutics, and Bitterroot. The other authors declare no competing interests. **Data and materials availability:** All data are available in the main text, supplementary materials, or online storage. Raw and processed mouse scRNA-seq data as well as processed human scRNA-seq data are available at the NCBI's Gene Expression Omnibus database under accession no. GSE224959. Raw human scRNA-seq data are deposited in the controlled access repository Data Use Oversight System (DUOS; https://duos.broadinstitute.org) under accession no. DUOS-000150 under the restrictions listed by this system. **License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq3061
Materials and Methods
Figs. S1 to S28
Tables S1 to S11
References (*51*–*67*)
Data S1 to S9
MDAR Reproducibility Checklist

View/request a protocol for this paper from *Bio-protocol*.

Submitted 31 March 2022; resubmitted 22 February 2023
Accepted 2 June 2023
10.1126/science.abq3061

By **Ananya Sen**

At my own pace

"Tell me your names and which year of graduate school you're in," the visiting speaker said. Our department was hosting them to give a guest seminar, and six of us graduate students had been chosen to accompany them for lunch. I was looking forward to talking science and learning about the speaker's career, but I dreaded this question. My discomfort grew as everyone else gave their answers. When my turn came, I swallowed and said I was in my sixth year. "I think if someone takes more than 5 years the student is the problem," the speaker replied tersely. I was devastated.

Right from the start of grad school, I had been aware that the requirements of my department and adviser might make my Ph.D. take longer than average. Per my department, students need to publish two first-author papers to graduate. My adviser requires students to take the lead on all experiments so they have full ownership of their work and develop an arsenal of scientific techniques. I knew this going in. Although I didn't relish the thought of getting by on a meager graduate student salary for longer than necessary, it wasn't a dealbreaker for me. I was comfortable with the potential trade-off of taking longer to finish in order to become an independent scientist and pursue the research that excited me. My adviser had a list of projects he wanted to tackle and I chose one to work on, knowing I could draw on his expertise and support from lab mates as I went.

"No one's Ph.D. journey is the same, and no one deserves to be shamed for how long it takes."

But as the years passed, I began to worry. I investigated 10 different hypotheses—just to prove each a dead end. We didn't want to publish papers that raised more questions than they answered. Nor did we want to abandon the project; as we eliminated each hypothesis, we felt we must be getting closer to the answer. So, each time the only option was to formulate the next likely hypothesis and start over in a different direction. But it seemed I was getting nowhere closer to graduating. The lack of tangible progress was wearing me down, and the raised eyebrows whenever I told anyone how long I had been in graduate school did nothing to lift my spirits.

The speaker's comment was the final straw. I went to my adviser and broke down. Didn't the speaker understand how hard it was to conduct research? Why were other scientists judging me for how long I had been in graduate school? Why did I always need to apologize for it?

My adviser was kind and supportive, as I had known he would be. He pointed out that over the past 6 years I had mastered many difficult techniques. My growing expertise meant no one could suggest an experiment I hadn't already done. If I didn't have the answer to my research question yet it was because the answer was complex, not because of personal shortcomings. My adviser had been working on this research problem for more than 3 decades, after all.

I was slow to stop blaming myself. But eventually, I saw what he was saying. I had done everything I could to answer my research question—running experiments in parallel to attack multiple hypotheses more efficiently, working late nights, and coming to lab before sunrise to set up experiments. I should be proud of my hard work and thorough investigation of many promising but ultimately unfruitful directions. I had nothing to apologize for.

My new perspective helped calm me when other students in my cohort graduated before I did. It helped me ignore the jabs and explain why my Ph.D. journey was taking longer. It helped me push forward with determination and confidence. And eventually, thanks to that hard work and persistence, and my adviser's continued guidance and support, we hit on a hypothesis that panned out. Seven-and-a-half years after starting my Ph.D., I graduated.

Toward the end, I realized many graduate students were in the same boat as I was, hiding in their labs and not talking to anyone just to avoid answering how long they had been in the program. I wish we hadn't been made to feel that way. I wish I could tell them—and those who made them feel bad—what I came to learn: No one's Ph.D. journey is the same, and no one deserves to be shamed for how long it takes. ■

Ananya Sen is a science writer for the Carl R. Woese Institute for Genomic Biology at the University of Illinois Urbana-Champaign. Send your career story to SciCareerEditor@aaas.org.

Ultra-Low Adherence Cell Culture Plate

SARSTEDT introduces the BIOFLOAT 96-well Ultra-Low Adherence cell culture plate specifically for culturing spheroids. The BIOFLOAT surface is anti-adhesive, which repels cells of all types, preventing them from attaching to the surface of the plate. This, along with the round bottoms of the wells, results in cells growing into 3D spheroids. The BIOFLOAT surface enables rapid spheroid formation—between 8 and 24 hours—with very high circularity and reproducibility compared to other surfaces. The chemical composition of the BIOFLOAT surface is very strong, which allows it to withstand multiple washing steps and/or mechanical action from pipetting. Additionally, the BIOFLOAT surface has proven to be very reliable, even when using difficult cells. In fact, many cell lines and some primary cells have already been successfully tested for spheroid formation on the BIOFLOAT surface.

SARSTEDT
For info: 1-800-257-5101
www.sarstedt.com

Tissue-Clearing Reagents for Plants

Tokyo Chemical Industry offers Tissue Clearing Reagent iTOMEI D and Tissue Clearing Reagent iTOMEI M, which enable transparency of plant organs and clear detection of fluorescent protein in 3D imaging and are applicable in many plant species, such as *Oryza sativa*, *Arabidopsis thaliana*, and *Marchantia polymorpha*. Tissue Clearing Reagent iTOMEI D is a decoloring solution to elute stains of plant organs like chlorophyll, and it enables transparency of plants by simple osmosis in just a few days. The main advantage of Tissue Clearing Reagent iTOMEI D is that it can suppress autofluorescence in plant organs while retaining the fluorescence of fluorescent proteins, such as GFP and tdTomato, coexpressed as reporter genes. Meanwhile, Tissue Clearing Reagent iTOMEI M is a mounting solution suitable for observation using silicon lens, which adjusts the refractive index in the plant organ to 1.40.

Tokyo Chemical Industry
For info: +81 (0)3 5640 8878
www.tcichemicals.com

FDA-Certified Tubing Solutions

Biotech Fluidics has launched a range of traceable, certified PEEK, and fluoropolymer-flexible tubing for use in regulated environments. In many medical and life science applications, biocompatibility requirements from regulatory standards including ISO 10993 (medical device and dental materials), US Pharmacopeia Class VI (toxicity), FDA title 21, and EU Regulation 1907/2006 (DEHP content) strictly govern what flexible plastic tubing can be used. To serve these medical and life science applications, Biotech Fluidics is now able to offer high quality, FDA-certified flexible plastic tubing in bulk and pre-cut quantities. In addition to regulatory compliance, every supply of this new tubing comes with traceability information including manufacturing lot number. If required, Biotech Fluidics can also deliver sterilized tubing assemblies prepared in a clean room environment.

Biotech Fluidics
For info: +1-612-703-5718
www.biotechfluidics.com/products/tubing

Microcentrifuge with Hydrocarbon Cooling

Eppendorf offers Centrifuge 5427 R—the first Eppendorf microcentrifuge with hydrocarbon cooling to contribute to an even more sustainable laboratory environment. With this offering, customers can now perform various molecular and cell biology applications while using a refrigerated device that contains a natural cooling agent with a GWP of almost zero. Natural cooling agents, like R290 (propane), have a comparably low global warming potential (GWP) to CO_2 (<3), while conventional refrigerants such as R134a have a GWP of 1430 and thus have a disproportionately greater impact on global warming when released into the environment. The ACT label certification of Centrifuge 5427 R makes it even easier for customers to choose a more sustainable product. Due to its compact footprint and the dual-row rotor FA-45-48-11 for up to 48 x 1.5/2 mL tubes, it is the optimal solution for laboratories with a high sample throughput.

Eppendorf
For info: +49 2232 418-0
https://eppendorf.group/5427R

2D-Barcoded Tube Whole Rack Reader

Ziath announces a revolution in cryogenic sample management—the Ri-Track Mirage 2D barcoded tube whole rack reader. Traditionally, identifying the identity of frost-covered racks of 2D barcoded tubes taken from low-temperature storage has been a time-consuming and error-prone process. Drawing upon proprietary Radio Frequency Identification (RFID) tagging technology, the Ri-Track Mirage can reliably read 2D barcoded racks covered in ice and with racks taken directly from vapour-phase liquid nitrogen storage. The need to warm your rack to thaw it out—which may degrade thermally sensitive samples—to read your identity of your sample rack is eliminated. Extensive testing has shown that Ri-Track RFID technology can withstand immersion and retrieval from vapour-phase liquid nitrogen more than 10,000 times without losing any stored data or showing any physical degradation. Ri-Track is therefore suitable for all forms of traceable long-term cryogenic sample storage.

Ziath
For info: +1-858-880-6920
www.ziath.com

Lab Balance

Weighing in a regulated environment can be complex, with stringent compliance directives and controls required to deliver the accuracy and reliability of results required by the pharmaceutical industry. One of the technologies helping scientists to achieve this level of accuracy is the Cubis II balance series from Sartorius. All weighing modules of the Cubis II balance are designed for intuitive operation, further aided by intelligent diagnostic systems. These design elements guarantee a higher degree of repeatability for different workflows while lowering the probability of human error during measurement. It offers modern user interfaces, pharmaceutical and GxP compliance, including data handling, data integrity, and connectivity, ergonomic sample handling, easy process integration, and unlimited communication at the highest level of accuracy and precision.

Sartorius
For info: +4-551-308-0
www.sartorius.com

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.